Job Market Signalling, Stereotype Threat, and Counter-Stereotypical Behaviour.

Richard Chisik*

October 2013

ABSTRACT: We introduce stereotype threat in a multiple-productivity signalling model. Existence of multiple self-fulfilling stereotypes, which can generate statistical discrimination, is more likely if there is less variance in the ability distribution. A low endogenously-correct stereotype about a group forces higher-ability group members to choose a higher-productivity and a higher separating signal, thereby engaging in counter-stereotypical behaviour. This counter-stereotypical behaviour causes the remaining partially-pooling group to have lower average productivity, reinforcing the negative stereotype. The co-existence of stereotype threat and counter-stereotypical behaviour can explain the simultaneity of lower wages and higher education attainment in a group facing labour-market discrimination.

JEL Classification: J70, D82, O15

* Department of Economics, JOR-211, Ryerson University, Toronto, ON, M5B 2K3.
E-mail: rchisik@arts.ryerson.edu. Phone: +1-416-979-5000 Ext. 4620.

Acknowledgements. I am grateful to several anonymous referees and the many helpful seminar participants at Ryerson University, the Canadian Economic Association 2012 meetings, and the Atlantic Canada Economic Association 2013 meetings for their useful comments and suggestions. I am thankful for financial support for this project granted by the Office of the Dean of Arts at Ryerson University. Any remaining errors are my own.

I. Introduction

Stereotype threat has been an important topic for psychology researchers since Rosenthal and Jacobson's (1968) seminal work "Pygmalion in the Classroom," which analyzes how completely random information about student ability told to teachers at the beginning of the school-year becomes self-fulfilling.[1] Although provocative, self-fulfilling stereotypes suggest a uniformity of experience within each group. There is, however, large within group variation that self-fulfilling stereotypes cannot explain by themselves. In this paper we develop a new model of statistical discrimination that allows a subset of a group to overcome a negative stereotype.

We are motivated by the following empirical observations. Canadian-born visible minorities earn less than white Canadians with similar education and experience: 18 per cent less for men and 3 percent less for women.[2] There are many possible reasons for these wage differences and if statistical discrimination was part of the explanation, then we should also expect that Canadian-born visible minorities would view education as less rewarding and obtain less of it (so that a lower productivity stereotype is, in fact, statistically correct). The opposite, however, is true for Canadian-born visible minorities at the top of the ability distribution: they are more likely than white Canadians to have a university degree.[3] A similar, and more detailed result, occurs in the United States. There, when controlling for ability as measured on the Armed Forces Qualification Test (AFQT), black women of all abilities but the lowest receive more education than similar white women. Black men of all abilities but the lowest and highest receive more education than similar white men (Lang and Manove, 2011). Still, the earnings of black men are lower than whites at intermediate education levels but not at high levels

---

[1] Eden (1992), and Dvir et al. (1995) show that the Pygmalion effect can exist in the workplace as well as the classroom. The self-fulfilling prophecy can also occur even if the teacher, or the boss, does not take any actions. For example, simply reminding test takers of their race before a test can affect results on standardized tests (Steele and Aronson, 1995). This stereotype threat is also demonstrated with respect to gender and entrepreneurship by Gupta et al. (2008), and gender and math performance by Good (2008) and Kiefer (2007).

[2] The wage gap for each Canadian-born group versus whites is 8 percent for those with Chinese parents, 13 percent for South Asian descendants, and 19 percent for blacks (Skuterud, 2010).

[3] Canadian-born holders of university degrees number 55 percent of Chinese-Canadians, 51 percent of those with South Asian parents, 25 percent of blacks, and 25 percent of whites (Canadian Census, 2006). The percentages of university degree holders for each remaining Canadian-born group in the 2006 census are 58 for Korean-Canadians, 54 for Filipino-Canadians, 47 for Japanese-Canadians, 43 for West-Asian-Canadians, 38 for Arab-Canadians, 36 for Southeast-Asian-Canadians, and 28 for Latin-American-Canadians. The number for non-visible minorities, which is made up of whites and aboriginals is 24 percent.

(Arcidiacono et al, 2010 and Lang and Manove, 2011).[4] The above examples suggest that although a group may suffer discrimination, a subset of that group may counter their stereotype by obtaining extra education. In this paper we start by providing intuitive conditions for self-fulfilling stereotypes to occur. We then show how counter-stereotypical behaviour can exist along with stereotype threat and how together they can explain the above empirical observations.[5]

We employ the following multiple productivity signalling model with an endogenous productivity choice. During his formative years, a worker makes choices that determine his lifetime productivity. Higher productivity choices are more costly and these idiosyncratic effort-cost differences depend on the worker's type. The type is determined by the worker's natural ability as well as his nurturing environment. The given type and the chosen productivity are unobservable to the firm when they hire the worker, however, the worker may use an additional observable signal (such as university education) to inform the firm of his chosen productivity. A higher productivity worker has a lower cost of the additional signal. In addition to his unobservable type the worker has an observable and unalterable marking that we call a label. The label refers to the worker's caste, ethnic origin, gender, or skin color.

The novel feature of our model is the endogenous productivity choice embedded in a signalling model. We analyze how this choice is affected by label-specific stereotypes that interact with the worker's ability to signal his productivity and we show that this heuristic process can become self-fulfilling. Although the label is not correlated with the unknown parameters in any fundamental way the incentive effects of the stereotype correctly correlates them in equilibrium. The label then provides statistical information to the firm who engages in statistical discrimination.

[4] A parallel result for women is impeded by selection issues, however, recent work by Salisbury(2012) on the marriage decisions of civil war widows suggests that the causality may go from labor market opportunities to the marriage market and not in the other direction.
[5] As Lang and Lehmann (2012) state "while one challenge is to explain earnings differentials between black and white men, there is an even greater challenge, which is to explain the simultaneous existence of wage differentials among relatively low-skill male workers and their possible absence among high-skill male workers". In particular, we answer this challenge and our model provides an explanation for these stylized facts.

We consider sequential equilibria that satisfy a restriction on beliefs and we call these credible equilibria. With three or more productivities the equilibrium dominance based intuitive criterion of Cho and Kreps (1987) and a strict dominance based refinement yield very similar results. The slightly less-restrictive dominance based refinement allows a more intuitive presentation (especially of the accompanying figures) and we begin our analysis with that refinement; we then consider the intuitive criterion in a later section. In particular, with more than two productivities both refinements do not rule out partial-pooling equilibria that Pareto dominate a fully separating equilibrium and we analyze both types of equilibria.[6] It is the juxtaposition between the pooling group in a partial-pooling equilibria, which may exhibit stereotype threat, and the strategic separating response to a very discriminatory belief that is at the heart of our paper.

If certain conditions are satisfied, then multiple credible pooling equilibria exist. As the required pooling signal increases the benefit to choosing higher productivity increases as well, therefore, each type will (weakly) increase their productivity and the average productivity will increase. Hence, for one label there exists a partial-pooling equilibrium with a low stereotype and a low pooling expenditure and for another label there exists a partial-pooling equilibrium with a high stereotype and a high pooling expenditure. The first contribution of our analysis is showing that these credible, self-fulfilling, statistical discrimination equilibria are more likely to occur if there is less variance in the distribution of types. Hence, stereotype threat is predicted to be more likely to occur when natural ability is less diverse. Although obtained in a particular framework we expect that these results hold more generally.

Our second contribution results from analyzing stereotype threat in a signalling model. Existing models of self-fulfilling statistical discrimination assume that productivity is imperfectly observed and potential revealing signals (such as university transcripts) are not observed. In signalling models, on the other hand, productivity is unobservable, but the potentially revealing signal is perfectly observed. We

[6] There are stronger refinements which leave only a unique separating equilibrium with any number of productivities. We are interested in both pooling and separating equilibrium and especially the comparison between them. A dominance refinement is, therefore, the simplest way to isolate the equilibria of interest. Furthermore, there are criticisms of these stronger refinements, and also of the intuitive criterion, (which we discuss in section III.B), however, none are directed at dominance refinements. Still, in section IV.D. we show that our results also obtain with the intuitive criterion.

use the information structure of the signalling model to show that stereotype threat can occur even when a potentially separating signal is present. Although signalling models generally have many pooling equilibria, the productivities are exogenous so that there is no relationship between the pooling signal and the expected productivity. In our model, on the other hand, the benefit to choosing a higher productivity is directly related to the pooling signal so that an equilibrium with a higher pooling signal also has a higher expected productivity. This monotonic relationship between the pooling signal and the expected productivity is unique to our model and is a result of the endogenous productivity choice.

A very low pooling stereotype implies a very low pooling wage so that a higher productivity agent would separate and break this low stereotype (and dominated) pooling equilibrium. A label that suffers from a very low pooling stereotype will have more of its high ability (low-cost-type) members separating. The following situation is then possible. A higher productivity worker with a moderate or high stereotype label will pool at a moderate or high level of the signal and the same productivity worker with a low stereotype label will engage in counter-stereotypical behaviour and separate himself with an even higher level of the signal. Furthermore, the good stereotype can generate complacency and a reputational Dutch disease effect, whereby a worker with a moderate stereotype label has less reason to distinguish himself and he remains content in the pooling equilibrium.[7] In particular, it is the strategic behaviour of the high productivity worker with a very discriminated against label that generates his high achievement. Identification of this counter-stereotypical behaviour is our third contribution.

The number of possible pooling productivities in any partial-pooling equilibrium is increasing in the stereotype. For very high stereotypes most of the productivities will pool (the highest separates under the intuitive criterion). As the stereotype drops additional (and successively lower, starting from the highest) productivities separate leaving only the lower productivities to pool. Hence, our model predicts that labels that expect lower pay at intermediate education levels would have a higher percentage of their

[7] The Dutch disease, or as it is sometimes called the natural resource curse, is where an initial comparative advantage in a natural resource may generate lower long-run growth. It is explained as both a monetary (currency appreciation) and a real (production shifting) phenomenon. The reputational aspect alluded to in this paper suggests that a high-ability person with a high-reputation label would have less incentive to differentiate themselves through high levels of education and individual achievement than they would if they were from a lower reputation label.

members completing higher (separating) levels of education, which helps explain the Canadian data. It also predicts that the highest productivity workers in all labels, who separate and are, therefore, perfectly revealed, would have similar wages, which helps explain the US data. These empirical predictions are our fourth, and most important, contribution.

II. Related Literature.

Our paper contributes to both the signalling and the statistical discrimination literatures. In Spence's (1973) seminal work on signalling firms may discriminate against women by requiring a much higher level of the separating expenditure for women than for men. In this case, women would pool (and obtain less education) while men separate. For example, a woman with seven PhDs could be believed to be low quality if separating beliefs require at least eight. To support this non-credible equilibrium, however, the firm must believe that low-quality women would play a strictly dominated strategy. Furthermore, these beliefs do not affect the proportion of high-quality women and, therefore, they are not truly self-fulfilling. An important contribution of our model is showing that statistical discrimination in a signalling model can occur, however, it requires agents to have an endogenous quality choice.[8]

The empirical analysis of Lange (2007) suggests that employers learn a lot in the first few years and that there is a modest upper bound on the signalling gain from an additional year of schooling.[9] This limitation on the value of signalling is refuted by Arcidiacono et al (2010). They parse Lange's data into differing education levels and show that a college degree reveals information so that employers learn very quickly about the productivity of these (separating) workers, however, they learn more slowly for those with less education. These empirical results coincide very well with the model in this paper. In studies using Canadian data Ferrer and Riddell (2002, 2008) also find evidence of the separating power of degree completion. An important theoretical explanation for the value of signalling is given by Bidner (2012)

[8] As opposed to counter-stereotypical behavior, Feltovich et al. (2002) consider counter-signaling. In their model, with three types and noisy additional information on quality, the middle type separates but the highest type's expected payoff is greater if they rely on the additional information and do not choose a separating signal level.

[9] An interpretation of the static signaling model is that the job offer encompasses the whole trajectory of future earnings collapsed into a single time period. In as much as the first job placement may affect the entire career path of the employee signaling would still have value irrespective of how fast the initial employer learns about productivity. Furthermore, unless the employer accurately conveys to the market what they have learned about their employee then the original employer's learning has no effect on additional offers that the employee may receive.

who shows that credentials can determine a worker's team assignment. Since team skills affect the individual worker's productivity measure, the signalling value of credentials may be under estimated.

The pioneering work on self-fulfilling discrimination in an economic environment is Arrow (1973). In his framework, imperfect observability about a worker's productivity leads employers to offer wages based in part on stereotypes about the worker's label. When the expected wage affects the productivity decision, negative or positive stereotypes can become self-fulfilling. Coate and Loury (1993) build on Arrow's model to study the effect of affirmative action policies in a statistical discrimination model of job assignment, which allows for the analysis of this policy's joint effect on worker incentives and employer beliefs.[10] One question that can be directed at these models is whether their results would still obtain if the worker had an additional non-noisy signal at their disposal. The results in this paper suggest that they could still obtain and it could, in fact, be the signal that generates the statistical discrimination for the pooling group, however, it may also generate the counter-stereotypical behaviour that we introduce here.

An alternative approach to statistical discrimination is explored by Phelps (1972), Aigner and Cain (1977), and Lundberg and Startz (1983) who assume that employers have less precise measurements of productivity for certain labels. In Lang and Manove (2011), firms see the worker's education and they receive a noisy signal about productivity. The signal is assumed to be noisier for blacks of intermediate education than it is for whites at the same level, however, the difference is assumed to disappear as education increases. In this way, blacks may suffer statistical discrimination and intermediate ability blacks get more education than whites of the same ability. In contrast to these models we do not assume any difference in observation between the groups. For example, in Lang and Manove (2011), the extra education chosen by blacks of middle ability is a direct result of the assumed imperfect information of their other skills. Whereas their chosen formulation fits the data on intermediate ability black men it does not, however, fit the education difference between high ability black and white women or the difference

[10] Fryer (2007) extends Arrow (1973) and Coate and Loury (1993) to a dynamic framework. His key insight is that the first period hiring decision truncates the distribution so that a hired worker from a low self-fulfilling stereotype label is of higher potential quality. Hence, there may be belief flipping so that a hired worker from the low stereotype label is more likely to receive a promotion in the second period.

in university degrees granted to Canadian-born visible minorities and whites. Our model, on the other hand, correctly predicts that upper intermediate and, especially, higher ability black American women and Canadian-born visible minorities would have higher rates of university degrees than whites. The difference in our model is that the increased education arises as a reaction to a bad stereotype rather than an assumption on imperfect observability and, in this way, is endogenized at a more primitive level.

Underlying our model is that stereotypes exist and that they are shared by potential employers as well as members of the stereotyped group. Some evidence on employer's views is provided by Bertrand and Mullainathan (2004) for the USA and Galarzo and Yamada (2012) for Peru. In both studies the authors found that otherwise identical job applications were less likely to result in an interview if the applicant had a black (for the USA) or indigenous (for Peru) sounding name. Wilson (1996) finds that 132 out of 179 Chicago employers expressed negative views of inner-city black men. This percentage was the same when the sample was restricted to black employers (12 out of 15 expressed the same negative views). In a study that combines both of the above findings Hanna and Linden (2012) arbitrarily assigned caste to exams that were to be graded for a competition in India. They found that the lower caste exams received lower scores, however, the negative bias was more pronounced in lower caste graders.

In the next section we develop our economic framework. In the fourth section we establish our main results. Conclusions are in the fifth section.

III. A Model

A. The Economic Environment

We consider the interaction between one worker and a single firm. The worker begins by observing their type, θ, and their label, i. After observing his type and label, the worker chooses his unobservable productivity, $q_k \in \{q_0,\ldots,q_N\}$ and a level of credentials, $s \in \Re^+$, which may enhance as well as signal his productivity choice. The firm observes the worker's credentials and label (but not the productivity or type) and makes a wage offer, $w \in \Re$.

Higher productivity requires an additional expenditure that depends on the worker's type. The type is the worker's private information and it is drawn from a continuously differentiable and commonly known log-concave distribution function $F(\theta)$ with a non-increasing density and a finite lower bound, $\underline{\theta}$.[11] The perfectly observable label, $i \in \{A, B, C, \ldots\}$, corresponds to the worker's race, or sex, or caste, etc. It is common knowledge that the label contains no direct information, in the sense that the distribution of types is independent of i: $F(\theta|i) = F(\theta)$ for all i. The label may, however, provide information about the worker's unobservable productivity choice. The common stereotype about a worker with label i choosing productivity q_k is denoted as $\Pr(q_k|i) = \rho_i^k$ and the vector of stereotypes for label i is $\boldsymbol{\rho}_i$. As we are considering a single worker we will drop the subscript from the stereotype and write ρ^k for the q_k stereotype, and $\boldsymbol{\rho}$ for the collection of stereotypes.

Our idea is the following. The worker's productivity is determined by choices the worker makes during his formative years and is partially related to his primary and secondary education. The cost of these choices is determined by the worker's ability, family, and neighborhood. These naturally vary and a given worker has superior information as to his true cost of lifetime productivity investments. Hence, we make the assumption that the worker has private information about his specific type. The worker may also gather perfectly observable credentials by engaging in tertiary, or university, education. The skills and habits developed by a high productivity worker reduces his effort cost of obtaining these credentials. The relationship between the desired credential and the productivity choice is given by the worker's disutility of effort or cost function. The payoff to the worker's chosen credential level, s, may depend on the stereotype, $\boldsymbol{\rho}$, and is captured by the expected wage function.

Consider the worker's effort cost function. We assume that the worker's type does not directly affect his credential cost and we write this function as additively separable $C(\theta, q, s) = I(\theta, q) + c(q, s)$. Although the functions are continuous in q the worker chooses a discrete q_k. The productivities are ordered so that $q_{k+1} - q_k = \Delta > 0$, for all k. The relationship between the worker's type and his cost of becoming higher productivity is described by the multiplicatively separable function $I(\theta, q) = g(\theta)d(q)$, which is increasing in both arguments and convex in q. The unobservable θ indexes the additional cost of

[11] A non-increasing density and a log concave distribution function are provided by the Pareto, exponential, uniform, and Zipf, as well as restricted versions of the Weibull, power, beta, gamma, chi-squared, truncated normal, and log-normal distributions. For more on log-concave distributions see Bagnoli and Bergstrom (2005). For more on the Pareto and Zipf distributions see Axtell (2001).

being a higher productivity worker, an activity that is more costly for higher types: $I_{\theta q}(\cdot,\cdot) > 0$.[12] In order to concentrate on the moral hazard in quality choice and avoid selection issues, we assume that $I(\theta, q_0) = 0$ for all θ (and in assumption 2 below we assume that the lowest productivity worker is valuable under certainty).[13] The effect of the productivity choice on the credential cost is given by the function $c(q, s)$. This cost is strictly increasing and weakly convex in the credential and the complementarities between s and q imply that $c_{sq}(\cdot,\cdot) < 0$. Although a higher productivity reduces the credential cost, this reduction is mitigated as productivity increases further, so that $c_q(\cdot, s) < 0 < c_{qq}(\cdot, s)$. These restrictions on the cost function are assumption 1:

$$I(\theta, q) = g(\theta)d(q),\ g_\theta(\cdot) > 0,\ d_q(\cdot) > 0 = d(q_0),\ d_{qq}(\cdot) \geq 0.$$

$$c_{qs}(\cdot,\cdot) < 0 < c_s(q,\cdot),\ 0 \leq c_{ss}(q,\cdot),\ c_q(\cdot, s) < 0 < c_{qq}(\cdot, s).\quad q_{k+1} - q_k = \Delta > 0, \text{ for all } k. \tag{1}$$

Assumption 1 captures the following idea. A lower type worker may choose higher productivity since it lowers the cost of university credentials. The firm may thus rationally infer that a higher credential is associated with higher unobservable productivity indicating that the worker is a lower type.

Next consider the firm's objective. The firm's revenue function is $v(q)$ and the firm values higher productivity: $v_q(\cdot) > 0$.[14] The firm's expected revenue is: $v^E(s, i) = E_q[v(q)] = \Sigma_k[b_k(s, i)\cdot v(q)]$. The variable $b_k(s, i) = Pr(q_k|i, s)$ is the firm's posterior belief that the worker is productivity q_k, conditional on its observation of the worker's label and credentials. The firm offers a wage (w) that the worker can accept or reject. The firm chooses the wage to maximize its expected profit function: $\pi = -(v^E - w)^2$. Finally, we assume $v(q_0) > c(q_0, 0) \geq 0$, so that the lowest productivity worker is valuable and hired.[15] These profit function restrictions are assumption 2.

---

[12] We follow the convention that a subscript refers to the partial derivative of the function with respect to the subscript: $\partial I(\cdot, q)/\partial\theta \equiv I_\theta(\cdot, q)$.

[13] Weaker restrictions can be made. If the expected wage for the lowest productivity worker is greater than the total cost for the highest type, then our results hold under the more general ordering: $I_\theta(\cdot, q_0) \geq 0$.

[14] As in the original Spence (1973) model, allowing for a non-dissipative signal (so that the credential adds value for the firm) does not alter any of our main results. In particular, if we were to assume that education is more than a credential, then it would not alter the worker's or the firm's incentives or any of the equilibrium outcomes.

[15] Our assumptions are satisfied by the following parameterization which was used for the accompanying figures: $v(q_k) = k + 1$, $c(q_k, s) = s^2/v(q_k)$, $I(\theta, q_k) = (\theta - \underline{\theta})(v(q_k) - 1)^a$, $a \geq 1$, and $F(\hat{\theta}) = 1 - (\underline{\theta}/\hat{\theta})^\lambda$, where $\underline{\theta} = 1$.

$v_q(\cdot) > 0$, $v(q_0) > c(q_0, 0) \geq 0$, $\pi = -(v^E - w)^2$. (2)

The firm observes the label and the signal, but not the worker's type or productivity choice. Hence, the wage is a function $w(s, i) = v^E(s, i)$.[16] Given his type and label the worker chooses q and then s to maximize his expected utility: $Eu(\theta, i) = Eu(\theta, q(\theta, i), s(\theta, i), i) = E[w(s, i)] - C(\theta, q, s)$.

B. Description of the Equilibrium

A pure strategy sequential equilibrium (Kreps and Wilson, 1982) is a collection of strategies and beliefs $\{s(\theta, i), q(\theta, i), w(s, i), b_k(s, i)\}$ which satisfy the following three conditions. Sequential rationality for the worker and for the firm requires that:

$$\{q(\theta, i), s(\theta, i)\} \in \text{argmax}_{q,s} E[w - C(\theta, q, s)] \quad \text{(E1w)}$$

$$w(s, i) \in \text{argmax}_w - (v^E(s, i) - w)^2. \quad \text{(E1f)}$$

In defining Bayesian beliefs note that all functions depend on i solely through the stereotypes in $\boldsymbol{\rho}$, therefore, we write $\boldsymbol{\rho}$ instead of i. Also note that θ affects s only through the choice of $q(\theta, \boldsymbol{\rho})$. Hence, we can write $s(q, \boldsymbol{\rho}) = s(q(\theta, \boldsymbol{\rho}), \boldsymbol{\rho}) = s(q, i) = \text{argmax}_s Eu(\theta, q, s, \boldsymbol{\rho})$ is the credential chosen by a productivity q worker when the equilibrium stereotypes are $\boldsymbol{\rho}$. Bayes-consistency of beliefs requires that:

$$\text{If } s(q_j, \boldsymbol{\rho}) = s(q_k, \boldsymbol{\rho}) \text{ for all } q_j, q_k, \text{ then } b_k[s(q_k, \boldsymbol{\rho}), \boldsymbol{\rho}] = \rho^k; \quad \text{(E2bp)}$$

$$\text{If } s(q_j, \boldsymbol{\rho}) \neq s(q_k, \boldsymbol{\rho}) \text{ for any } q_j, q_k, \text{ then } b_k[s(q_k, \boldsymbol{\rho}), \boldsymbol{\rho}] = 1. \quad \text{(E2bs)}$$

Fully pooling equilibrium beliefs are described by (E2bp), where each productivity worker chooses the same credential and the firm learns nothing from the choice. In a separating equilibrium (E2bs) each productivity worker chooses differing credentials and the posterior beliefs recognize this separation. We are also interested in partial pooling equilibrium, whereby a strict subset of the

[16] We could also assume that there are at least two firms who bid for the worker and we would obtain all of the same results, however, we would need to modify the model so that there is more than one receiver of the signal. Either formulation ensures that the worker has monopoly power. Giving the worker monopoly power yields a straightforward division of the gains from trade. Other specifications are, of course, possible, however, as long as they allow the worker to retain some of the surplus he creates by being high productivity, these alternative specifications do not change our results. Furthermore, by assuming that the firm bids for the worker we rule out the possibility of price or wage signaling by the worker. We do not make this assumption to detract from the importance of price signaling, but rather to limit our analysis to a one dimensional signal.

productivities pool and the remaining ones separate. In the case where q_0 through q_J pool (where $J < N$) and q_{J+1} through q_N separate this condition can be written as:

If $s(q_j, \boldsymbol{\rho}) = s(q_J, \boldsymbol{\rho})$ for all $j \leq J$, and if, for all $t \in [1,\ldots, N - J]$, $s(q_{J+t}, \boldsymbol{\rho}) \neq s(q_k, \boldsymbol{\rho})$ for all $k \neq J + t$, then $b_{J+t}[s(q_{J+t}, \boldsymbol{\rho}), \boldsymbol{\rho}] = 1$ and $b_j[s(q_j, \boldsymbol{\rho}), \boldsymbol{\rho}] = \rho^j$. (E2bpp)

As is well known, a sequential equilibrium does not place adequate structure on beliefs off of the equilibrium path. We follow the existing literature in refining the set of sequential equilibrium to only include those that have reasonable beliefs. Although there is no strict consensus on what constitutes reasonable beliefs we use the least contentious and least restrictive refinement. In particular, we restrict the set of sequential equilibrium beliefs to only include those that place zero probability on the event that a worker played a strictly dominated strategy. We use the term credible to describe beliefs and to describe equilibria that satisfy this dominance refinement.

A strategy s is strictly dominated by s' if $\underset{w'}{Min}\, Eu(\theta, q, s', \boldsymbol{\rho}, w') > \underset{w}{Max}\, Eu(\theta, q, s, \boldsymbol{\rho}, w)$. Unfortunately, there is no upper or lower bound on the wage that the firm could offer, therefore, no strategy is strictly dominated for any productivity of worker for some belief on the firm's offered wage. Still, sequential rationality for the firm implies that the bounds on the wage they would offer in response to any signal level are given by the full information values. In particular, if we restrict the worker to beliefs such that the expected wage $w(s, \boldsymbol{\rho}) \in W^* = [v(q_0), v(q_N)]$, we can then define the set of productivities, $\xi_q(s)$, for which strategy s is not strictly dominated.[17] Our dominance refinement requires that after seeing some strategy, s'', the firm puts probability on productivity q_k if and only if q_k is an element of this set of un-dominated strategies: $q_k \in \xi_q(s'')$. Put concisely:

Define $\xi_q(s) = \{q(\theta, \boldsymbol{\rho})|\, \sim\exists\, s'$ satisfying $\underset{w' \in W^*}{Min}\, Eu(\theta, q, s', \boldsymbol{\rho}, w') > \underset{w \in W^*}{Max}\, Eu(\theta, q, s, \boldsymbol{\rho}, w)\}$

Then, $Pr[q_k(\theta, \boldsymbol{\rho})|s''] > 0$ if and only if $q_k \in \xi_q(s'')$ (E2bd)

[17] Our refinement and its exposition is based on the section entitled "Domination-Based Refinements of Beliefs" in Mas-Colell et al (1995, pp. 468-470). If instead, as in Mas-Colell et al (1995), we assume that two or more firms compete in Bertrand fashion in hiring the worker, then we would need to modify the set W^* to only include possible equilibrium responses by the firms. This is because, given an arbitrary belief about the other firm's wage offer, no wage offer is strictly dominated for a firm. Hence, the assumption of one firm with a quadratic loss payoff function mildly simplifies the exposition of this part of the model and has no effect in the remainder.

Note that we include the posterior belief (through its effect on the firm's action) in condition (E2bd).

The dominance refinement we use is common in the literature. A condition like E2bd can be found in Cho and Kreps (1987, pp. 199-201), Fudenberg and Tirole (1991, p. 330), and Mas-Colell et al (1995, pp. 468-470) among others. Some of the literature on signalling games utilizes more powerful belief refinements such as the intuitive criterion of Cho and Kreps (1987), or their stronger D1 which is very similar to the divine equilibrium of Banks and Sobel (1987). These refinements are motivated by the attempt to apply Kohlberg and Merten's (1986) strategic stability concept to signalling games, and none of these stronger ones are universally accepted, even by these original researchers. For example Cho and Kreps (1987, pp. 202-203) state that "Despite the name we have given it, the Intuitive Criterion is not completely intuitive. (It is certainly less intuitive than applications of dominance.)" They then proceed to discuss (p.203) a verbal critique that Joe Stiglitz gave to their proposed refinement.[18] In their conclusion (p.215) they continue to discuss the limitation of these stronger refinements when they say, "we do not mean to advocate all the tests we have described…the tests we have devised are very powerful in applications; perhaps too powerful." For their part, Banks and Sobel (1987) do not attempt to provide economic intuition for their divine equilibrium. Mailath et al. (1993) also express reservations with forward induction refinements, especially when they refine away pooling equilibria that Pareto dominate the separating equilibrium; therefore, they introduce undefeated equilibrium.[19] They also argue that a single solution concept that applies to all games does not exist and that different refinements could be employed in the analysis of a single game (p.265).

---

[18] This critique is described for the beer-quiche example, however, it can be easily applied to a labor market signaling game as follows. When breaking a candidate pooling equilibrium the intuitive criterion looks for a deviation from the pooling equilibrium that the high quality would choose if believed to be high quality and that the low quality would not mimic. The single-crossing property guarantees that such a deviation exists. The problem is that the low quality does not mimic this deviation because they compare it to their payoff at the proposed but no longer valid pooling equilibrium. If they do not mimic this deviation, however, they no longer receive the pooling equilibrium payoff and instead receive the low quality separating payoff. Being as the refinement presumes forward induction, the low quality should foresee this outcome and mimic this deviation. Hence, such a deviation cannot be considered as revealing the high quality and it would not break the proposed pooling equilibrium.

[19] The perfect sequential equilibria of Grossman and Perry (1986) can also be used to justify a deviation to a Pareto dominating pooling equilibrium. Unfortunately, their credible belief restriction also eliminates all pooling equilibria; therefore, even in simple signaling models, if the percentage of high quality types is large (so that pooling equilibria Pareto dominate the separating equilibria), then a perfect sequential equilibrium may fail to exist.

Our choice of the weaker dominance refinement is not only because it is more intuitive and less contentious than the stronger intuitive criterion (let alone D1 or divinity), but also because with three or more productivities the intuitive criterion would not remove the multiple pooling equilibria that we describe here (with only two productivities the intuitive criterion refines away all but the lowest cost separating equilibrium).[20] In the penultimate section we analyze our model with the further restriction of the intuitive criterion and we show that it does not alter our results. For all of these reasons we proceed with the more intuitive dominance refinement.

In defining the third equilibrium condition note that since the wage $w(s(q, \boldsymbol{\rho}), \boldsymbol{\rho}) = v^E(s(q, \boldsymbol{\rho}), \boldsymbol{\rho})$ is a function of the stereotype, the productivity decision is also a function of $\boldsymbol{\rho}$. Next note that if $q_k(\theta, \boldsymbol{\rho})$ is the optimal productivity choice for some type θ when stereotypes are $\boldsymbol{\rho}$, then the highest type choosing q_k, which we denote as θ^k, is the solution to

$$Eu[\theta^k, q_k, s(q_k, \boldsymbol{\rho}), \boldsymbol{\rho}] - Eu[\theta^k, q_{k-1}, s(q_{k-1}, \boldsymbol{\rho}), \boldsymbol{\rho}] = 0. \qquad (3)$$

The two single crossing properties given in assumption 1 ensure that for each q_k there is a unique $\theta^k(\boldsymbol{\rho})$ who is indifferent between productivity q_k and q_{k-1} and is, therefore, the highest θ choosing productivity q_k. All $\theta > \theta^k$ choose a lower productivity and all $\theta \leq \theta^k$ choose q_k or higher. Similarly all $\theta \leq \theta^{k+1}$ choose q_{k+1} or higher. Hence, $F(\theta^k(\boldsymbol{\rho})) - F(\theta^{k+1}(\boldsymbol{\rho}))$ is the probability that a worker with a given label i chooses productivity q_k. Now consider the stereotype that a worker with label i chooses an arbitrary q_k over any other choice $q_j \neq q_k$. This probability is given by

$$\rho^k = Pr(q = q_k|i) = Pr(Eu[\theta, q_k, s(q_k, \boldsymbol{\rho}), \boldsymbol{\rho}] \geq Eu[\theta, q_j, s(q_j, \boldsymbol{\rho}), \boldsymbol{\rho}]),$$

which by assumptions (1) and (2) can be written as $Pr(Eu[\theta, q_k, s(q_k, \boldsymbol{\rho}), \boldsymbol{\rho}] \geq Eu[\theta, q_{k+1}, s(q_{k+1}, \boldsymbol{\rho}), \boldsymbol{\rho}])$ and $Pr(Eu[\theta, q_k, s(q_k, \boldsymbol{\rho}), \boldsymbol{\rho}] \geq Eu[\theta, q_{k-1}, s(q_{k-1}, \boldsymbol{\rho}), \boldsymbol{\rho}])$. Using equation (3) these can be written as:

$Pr(Eu[\theta, q_{k+1}, s(q_{k+1}, \boldsymbol{\rho}), \boldsymbol{\rho}] < Eu[\theta^{k+1}, q_{k+1}, s(q_{k+1}, \boldsymbol{\rho}), \boldsymbol{\rho}])$ and

$Pr(Eu[\theta, q_k, s(q_k, \boldsymbol{\rho}), \boldsymbol{\rho}] \geq Eu[\theta^k, q_k, s(q_k, \boldsymbol{\rho}), \boldsymbol{\rho}])$, which is equivalent to

[20] The D1 or divine equilibrium concept would eradicate all pooling equilibrium with any number of productivities, but as Cho and Kreps (1987, p.215) say, "We ourselves find the D1 test very strong in the context of the Spence model".

$Pr[C(\theta, q_{k+1}, s(q_{k+1}, \boldsymbol{\rho})) > C(\theta^{k+1}(\boldsymbol{\rho}), q_{k+1}, s(q_{k+1}, \boldsymbol{\rho}))]$ and $Pr[C(\theta, q_k, s(q_k, \boldsymbol{\rho})) \leq C(\theta^k(\boldsymbol{\rho}), q_k, s(q_k, \boldsymbol{\rho}))]$, or $Pr(\theta^{k+1}(\boldsymbol{\rho}) < \theta)$ and $Pr(\theta \leq \theta^k(\boldsymbol{\rho}))$ or $Pr(\theta^{k+1}(\boldsymbol{\rho}) < \theta \leq \theta^k(\boldsymbol{\rho})) = F(\theta^k(\boldsymbol{\rho})) - F(\theta^{k+1}(\boldsymbol{\rho}))$.

Proceeding in a similar fashion, we can write the probability of the lowest and highest productivity choices, q_0 and q_N. We note that θ^1 and θ^N are the lowest and highest critical types in our model, therefore, with some abuse of notation, we can consider θ^0 as equal to the upper bound of the distribution and θ^{N+1} as equal to $\underline{\theta}$. For notational simplicity we, therefore, write $F(\theta^{N+1}(\boldsymbol{\rho})) = 0$ and $F(\theta^0(\boldsymbol{\rho})) = 1$. The third equilibrium condition, consistency with a common prior distribution, can then be written as follows:

$$\rho^k = F(\theta^k(\boldsymbol{\rho})) - F(\theta^{k+1}(\boldsymbol{\rho})). \tag{E3c}$$

Hence, stereotypes about a particular label's productivity choices induce the same measure of workers with that label to choose each productivity. Consistency, in this context, is essentially a fixed point argument. That is, the stereotype is an interim belief on productivity choices that is self-fulfilling and is also consistent with the commonly known prior distribution of types.

Our signalling model differs from the traditional approach in the following respect. The worker chooses his unobservable productivity as well as his observable signal. The common prior on the distribution of types is, therefore, not over productivity but rather over the investment costs necessary in the productivity choice. If the worker could not choose his productivity, then there would be one correct belief about the probability of each productivity in any credible separating or pooling equilibrium. Consistency with the common prior is straightforward in the standard model. In our model it is the formal statement of this condition that shows us how stereotype threat may arise in a signalling model.

IV. The Main Results

A. Separating Equilibrium.

We start by analyzing the unique credible separating equilibrium for this economic environment. As the description of this lowest-cost separating, or Riley (1979), equilibrium has been well-described in the literature, we are brief. Still, it introduces our method of endogenizing the measure of each productivity. Furthermore, the payoffs in the credible separating equilibrium affect the existence of multiple credible pooling equilibria and are necessary for the following sections.

Figure 1 goes about here.

The results of this section are illustrated in figure 1 for the three productivity case: $q = \{l, m, h\}$. We see there a wage function for beliefs that support the separating equilibrium. The indifference curve for the low productivity worker is shown at his full-information utility level. Where this low indifference curve crosses the full information wage for the medium productivity worker yields the separating signal, s^{sm}. We draw the medium productivity indifference curve through s^{sm} and where this medium indifference curve crosses the full information wage for the high productivity worker yields the separating signal, s^{sh}. The high productivity indifference curve is shown going through this highest signal level and from the single-crossing property in assumption 1 it lies above the medium productivity indifference curve at $s = 0$, which in turn lies above the low productivity indifference curve when it crosses the vertical axis.[21] The difference in utility levels (as measured on the w axis) yields an increasing (and invertible) function of the critical types for the separating equilibrium: θ^{sh} and θ^{sm}. A worker with low costs ($\theta \leq \theta^{sh}$) becomes high productivity, one with medium costs($\theta^{sh} < \theta \leq \theta^{sm}$) becomes medium productivity, and one with high costs ($\theta > \theta^{sm}$) becomes low productivity.

In describing a separating equilibrium it is useful to write the worker's expected utility as a function of the posterior belief: $Eu(\theta, q, s, \boldsymbol{\rho}, b_k(s, \boldsymbol{\rho})) = Eu(\theta, q, s, \boldsymbol{\rho})$. The full-information signals for each productivity are $s^{k*} = 0$ and the separating equilibrium signals are s^{sk}. In a separating equilibrium a lower productivity worker does not mimic the next highest productivity worker's signal, therefore, the lowest productivity worker prefers to reveal himself with the full information signal so that $s^{s0} = s^{0*}$. The highest signal, s^{sk+1}, that a q_k worker would copy, even if believed to be q_{k+1}, is given by:

$$Eu[\theta, q_k, s^{sk+1}, \boldsymbol{\rho}, b_{k+1} = 1] = Eu[\theta, q_k, s^{sk}, \boldsymbol{\rho}, b_k = 1], \tag{4}$$

To simplify the exposition we make the following assumption 3, which says that a positive measure of types choose each productivity.

For each q_k there exists a $\theta(q_k) > \underline{\theta}$ such that

$$Eu[\theta(q_k), q_{k+1}, s^{sk+1}, \boldsymbol{\rho}, b_{k+1} = 1] < Eu[\theta(q_k), q_k, s^{sk}, \boldsymbol{\rho}, b_k = 1] > Eu[\theta(q_k), q_{k-1}, s^{sk-1}, \boldsymbol{\rho}, b_{k-1} = 1]. \tag{5}$$

[21] Ignoring the investment cost and rearranging the expected utility yields the equation for each indifference curve as $w = \bar{u} + c(q, s)$.

We can now show the following (the proof is in the appendix).

Proposition 1: *If assumptions 1 – 3 are satisfied, then there exists a unique credible separating equilibrium where all $\theta \leq \theta^{sN}$ choose $\{q_N, s^{sN}\}$, all $\theta \in (\theta^{sk+1}, \theta^{sk}]$ choose $\{q_k, s^{sk}\}$, and all $\theta > \theta^{s1}$ choose $\{q_0, s^{0*}\}$. Furthermore, $\rho^k = F(\theta^{sk}(\boldsymbol{\rho})) - F(\theta^{sk+1}(\boldsymbol{\rho}))$ for all q_k.*

B. Stereotype Threat

We now consider fully pooling equilibria. In a fully pooling equilibrium all productivities choose the same pooling credential, s^p, therefore, from (E2bp) the posterior beliefs are the same as the prior beliefs and the expected wage is the same for each productivity. The key is that the prior belief is determined endogenously by (E3c) and this equilibrium correspondence may have multiple solutions.

Figure 2 goes about here.

In figure 2 we consider only two productivities in order to simplify the graphical presentation. We show the pooling wage function $w^p(s^p, \boldsymbol{\rho})$ as increasing in s^p, which is confirmed for any finite number of productivities in the proof of proposition 2. On this wage function we show two pooling equilibria. The good pooling equilibrium requires a higher university education level than does the bad one: $s^{pg} > s^{pb}$. In both equilibria the utility level (as measured along the wage axis) is greater for the high productivity worker, but, crucially, this difference is larger for the good equilibrium (because of the single-crossing-property and because the good equilibrium indifference curves cross the wage function further from the wage axis). Hence, when comparing different pooling equilibria we see that a higher pooling signal not only corresponds to a higher wage but also a better stereotype: $\theta^{hg} > \theta^{hb}$, so that $\rho^{hg} > \rho^{hb}$. In the N-productivity case, we also show that each θ^k increases when the pooling credential increases, and that ρ^N increases and that ρ^0 decreases, however it would not make sense to say that each ρ^k increases (because they need to add up to one). Instead we speak of the group average stereotype and we show that it is better in a pooling equilibrium with a larger pooling credential. The group average stereotype or expected productivity for a label with a vector of stereotypes $\boldsymbol{\rho}$ is given by:

$$\mathrm{E}q(\boldsymbol{\rho}) = \sum_{k=0}^{N} \rho^k \cdot q_k \tag{6}$$

Proposition 2: *(i) If assumptions 1 – 3 are satisfied, then there exists a continuum of fully pooling sequential equilibria and in each equilibrium each type's productivity choice is weakly increasing in, and uniquely determined by, the pooling expenditure. In a pooling sequential equilibrium with a larger pooling expenditure: (ii) the group average stereotype is higher; (iii) the probability of the highest productivity increases and the probability of the lowest productivity decreases; (iv) the pooling wage is higher.*

Proof: (i) For a pooling equilibrium equation (3) is written as: $Eu[\theta^k, q_k, s^p, \boldsymbol{\rho}] = Eu[\theta^k, q_{k-1}, s^p, \boldsymbol{\rho}]$. The expected wage is the same for any worker choosing s^p, therefore, this equation can be rewritten as:

$$\mathrm{I}(\theta^k, q_k) - \mathrm{I}(\theta^k, q_{k-1}) = g(\theta^k)[d(q_k) - d(q_{k-1})] = c(q_{k-1}, s^p) - c(q_k, s^p). \tag{7}$$

Now, $g(\theta)$ is continuous and strictly increasing, therefore, it has a continuous and strictly increasing inverse function and the critical type for each quality is given as:

$$\theta^k = \delta(s^p, q_k) = g^{-1}\left(\frac{c\left(q_{k-1}, s^p\right) - c\left(q_k, s^p\right)}{d\left(q_k\right) - d\left(q_{k-1}\right)}\right). \tag{8}$$

We simplify notation by writing $\delta(s^p, q_k) = g^{-1}(s^p, q_k, q_{k-1})$. From assumption 1, $c(q_{k-1}, s^p) - c(q_k, s^p)$ is positive, continuous, and increasing in s^p, therefore, $\delta(s^p, q_k)$ is positive, continuous and increasing in s^p so that the critical θ^k for each q_k is increasing in s^p. Hence, in an equilibrium with a higher s^p each type either increases their productivity or stays the same.

(ii) From assumption 1, $c(q_{k-1}, s^p) - c(q_k, s^p)$ is positive and decreasing as productivity further increases and $d(q_k) - d(q_{k-1})$ is positive and non-decreasing in further productivity increases, therefore, $\delta(s^p, q_k) > \delta(s^p, q_{k+1})$. Hence, because $F(\theta^k)$ is increasing and continuous we have that

$$\rho^k = F(\theta^k) - F(\theta^{k+1}) = F(\delta(s^p, q_k)) - F(\delta(s^p, q_{k+1})) > 0. \tag{9}$$

Remember that $F(\theta^0) = 1$ and $F(\theta^{N+1}) = 0$. Substituting equations (7), (8), and (9) into equation (6) yields

$$\mathrm{E}q(s^p, \boldsymbol{\rho}) = \sum_{k=0}^{N} \rho^k \cdot q_k = \sum_{k=0}^{N} [F(\theta^k) - F(\theta^{k+1})]q_k = q_0 + \sum_{k=0}^{N-1} F(\delta(s^p, q_{k+1}))[q_{k+1} - q_k] \tag{10}$$

and its derivative with respect to s^p is given as

$$\frac{\partial Eq(\cdot, \boldsymbol{\rho})}{\partial s^p} = \sum_{k=0}^{N-1} F_\theta(\theta^{k+1}) \cdot \delta_s(\cdot, q_{k+1}) \cdot [q_{k+1} - q_k] > 0. \tag{11}$$

(iii) Differentiating (9) with respect to s^p we have:

$$\frac{\partial \rho^k}{\partial s^p} = F_\theta(\theta^k)\delta_s(\cdot, q_k) - F_\theta(\theta^{k+1})\delta_s(\cdot, q_{k+1}). \tag{12}$$

For ρ^0, equation (9) is $\rho^0 = 1 - F(\theta^1)$, so that (12) becomes $0 - F_\theta(\theta^1)\delta_s(\cdot, q_1) < 0$ and for ρ^N, equation (9) is $\rho^N = F(\theta^N)$, so that (12) becomes $F_\theta(\theta^N)\delta_s(\cdot, q_N) - 0 > 0$.

(iv) The wage in a pooling equilibrium is:

$$w^p(s^p, \boldsymbol{\rho}) = \sum_{k=0}^{N} \rho^k \cdot v(q_k) = \sum_{k=0}^{N} [F(\theta^k) - F(\theta^{k+1})]v(q_k) = v(q_0) + \sum_{k=0}^{N-1} F(\theta^{k+1})[v(q_{k+1}) - v(q_k)]. \tag{13}$$

Its derivative with respect to s^p is:

$$\frac{\partial w^p(s^p, \boldsymbol{\rho})}{\partial s^p} = \sum_{k=0}^{N-1} F_\theta(\theta^{k+1})\delta_s(\cdot, q_{k+1})[v(q_{k+1}) - v(q_k)] > 0. \tag{14}$$

The derivative in equation (14) is strictly positive, therefore, $w^p(s^p, \boldsymbol{\rho})$ is strictly increasing in s^p. □

In proposition 2 we develop the relationship between the group average stereotype, the pooling wage and the signalling expenditure. The result occurs because a higher pooling expenditure increases the signalling cost of a lower productivity and generates a movement to higher productivity. The monotonic relationship between the pooling wage and the pooling expenditure is an important result of our model and it does not occur in a signalling model where a worker's productivity is fixed.[22]

Although proposition 2 considers fully pooling equilibria it is straightforward to extend the results to partial pooling equilibria. In the type of partial pooling equilibrium we consider here all productivities q_0 through $q_J < q_N$ pool together and q_{J+1} through q_N separate. If we then replace productivity q_N with q_J in the proof of proposition 2 we would obtain the same result for a partial pooling equilibrium for productivities q_0 through q_J. The separating credentials for productivities q_{J+1} through q_N are as given in proposition 1, with the distinction being that the separating credential for productivity q_{J+1} is given by the indifference condition for productivity q_J who can either choose the pooling credential for

[22] Although we can show that the group stereotype and the stereotype for the highest productive must increase in an equilibrium with a higher pooling credential we do not say anything definite about each intermediate productivity stereotypes. Still, if the distribution of types is uniform and $c_{qqs}(\cdot, \cdot) > 0$ (as given in the example in footnote 15, then the result in part (iii) of the proposition can be extended to show that when s^p increases only ρ^0 decreases and all the remaining ρ^k increase.

the J productivities, s^{pJ}, or separate at s^{sJ+1}. Similarly, the separating credential for q_{J+2} is determined in relationship to that for q_{J+1}. We state these extensions as proposition 3. (The proof is in the appendix.)

Proposition 3: *There exists a continuum of partial pooling sequential equilibria for each $J < N$, such that productivities q_0 through q_J pool together at s^{pJ} and productivities q_{J+1} through q_N separate. The partial pooling wage, $w(s^{pJ}, \boldsymbol{\rho})$, and the group average stereotype are strictly increasing in the pooling expenditure, s^{pJ}. The highest pooling productivity, q_J,* is *indifferent between receiving $w^p(s^{pJ}, \boldsymbol{\rho})$ with an expenditure of s^{pJ} or receiving the full information wage, and making the separating expenditure, s^{sJ+1}, for productivity q_{J+1}. The separating expenditures for each productivity q_{J+2} through q_N are then chosen with respect to this separating expenditure for q_{J+1}.*

Although propositions 2 and 3 suggest that there exists a plethora of fully and partial pooling equilibria we need to consider if these beliefs satisfy (E2bd) and are, therefore, credible. For example, if $s^p = 0$, then all types choose productivity q_0 and, therefore, $\rho^0 = b_0(0, \boldsymbol{\rho}) = 1$. If, however, the investment cost is low for at least one type (as given by assumption 3), then that type would deviate to a separating credential and $s^p = 0$ could not be part of a credible pooling equilibrium.[23] More generally, for any proposed pooling belief, if any productivity prefers his separating equilibrium payoff to that at the proposed pooling equilibrium then he would break the pooling equilibria. Only pooling sequential equilibria that are preferred by all pooling productivities are credible. Whereas a credible separating equilibrium always exists, the same is not necessarily true of a credible pooling equilibrium.

Figure 3 goes about here.

To help understand when a credible pooling equilibrium exists refer to figure 3 where the indifference curves for the separating equilibrium as well as two different wage functions for a pooling equilibrium are shown. A credible pooling equilibrium must lie in the shaded area. Only one of the shown pooling wage functions can support a credible pooling equilibrium and a credible pooling equilibrium is shown on this wage function. The lower pooling wage function lies below the shaded area,

[23] From this point on when we speak generally of fully and partial pooling equilibria we refer to them both as pooling equilibria and use the simpler notation s^p to denote the pooling credential.

therefore, the high productivity worker would break a pooling equilibrium on this lower wage function. Hence, with two productivities existence of a credible pooling equilibrium requires that $w^p(s^p, \boldsymbol{\rho})$ increases rapidly in s^p.

The intuition for existence of a credible pooling equilibrium with several possible productivities is similar but contains an additional consideration. To develop it consider figure 4. First note that for low values of s^p a higher productivity separating indifference curves lie above that for a lower productivity, therefore, a steeper pooling wage function in this range not only permits existence of a credible pooling equilibrium but also increases the number of potentially pooling productivities. On the other hand as s^p is further increased beyond a critical value, denoted as $\bar{s}^c(l,m,h)$, we see that it is the lowest productivity who would gain the most from separating. Still, although the lowest productivity may separate for $s^p >$ $\bar{s}^c(l,m,h)$, he won't separate if the pooling wage is large enough. Hence, if the wage is increasing rapidly enough in the pooling expenditure, then the range of s^p that can be part of a credible pooling equilibrium is larger.

<u>Figure 4 goes about here.</u>

The three determinants of the expenditure on the pooling wage function are given in equation (14) in the proof of proposition 2. First, if the sorting condition is more pronounced, then the signal cost difference and, therefore, $\delta(s^p, q_k)$, is steeper, and more types choose to become higher productivity which, in turn, increases the wage. Second, if the distribution function increases rapidly, then the wage function will as well. For example, if most of the types are similar, with similar investment costs, then once the lowest type becomes high productivity the rest will quickly follow. Hence, it appears that if the variance of types is small then the wage function will increase rapidly in s^p and a credible pooling equilibrium exists. This last idea is verified below for the exponential, Pareto, and uniform distributions. Third, if the investment cost differential for different types is small, then the variance in costs, as opposed to types, is small and the wage function increases rapidly in s^p.

Whereas the sorting condition also affects the separating equilibrium indifference curves, the shape of the distribution and investment cost function only affect the pooling wage, therefore, we concentrate on these two functions in looking for sufficient conditions for a credible pooling equilibrium to exist. To help us understand the shape of the distribution function, define

$$\sigma(\theta) \equiv (\ln(F'))' = F''/F' \tag{15}$$

as the degree of concavity of the distribution. For two distributions F_1 and F_2 of the same type, with corresponding σ_1 and σ_2, if $\sigma_1(\theta) < \sigma_2(\theta) \leq 0$ for every θ, then distribution F_1 is relatively more concave than (and also first-order stochastically dominates) F_2. If $\sigma(\theta) \leq 0$, then the density function is non-increasing. The density (F') is positive, therefore, if it is non-increasing, then $(\ln(F))'' = \frac{FF''-F'F'}{F^2} \leq 0$, so that $\sigma(\theta) \leq 0$ is sufficient for the distribution function to be log concave. Formalizing the discussion of the previous two paragraphs we now show the following. (The proof is in the appendix.)

Proposition 4: *If assumptions 1 – 3 are satisfied, then:* (*i*) *there is a finite* $\lambda^* \geq 0$ *such that if* $\sigma(\underline{\theta}) \leq -\lambda^*$*, then a credible pooling equilibrium exists;* (*ii*) *the most pessimistic group average stereotype,* $Eq(\boldsymbol{\rho}) = q_0$*, is never part of a credible pooling equilibrium;* (*iii*) *a credible pooling equilibrium is more likely to exist, and the measure of credible pooling equilibrium expenditures for any set of (partially or fully) pooling productivities is larger, if* $\sigma(\theta)$ *is smaller; (iv) the required* λ^* *is smaller if* $I_\theta(\cdot, q)$ *is smaller.*

An important corollary of proposition 4 is that if a credible pooling equilibrium exists for some s^p, then credible pooling equilibria exist in the neighborhood of this s^p. This result is immediate because of the continuity of the wage function (which is given by the assumed continuity of the distribution, value and cost functions). Hence, combining this multiplicity of pooling equilibria with the results of proposition 2, which provides a rationale for ranking pooling equilibria, allows us to speak of self-fulfilling statistical discrimination, or stereotype threat, in labor market signalling models. [24]

Corollary 1: *Credible self-fulfilling statistical discrimination equilibria exist in signalling models with an endogenous productivity choice.*

[24] Although one pooling equilibrium may Pareto dominate another pooling equilibrium they are both credible if they both Pareto dominate the separating equilibrium. This occurs because when starting from one equilibrium, the firm may have beliefs about other pooling education levels so that neither productivity (nor underlying type) is playing a strictly dominated strategy in the original Pareto dominated pooling equilibrium.

A second important corollary of proposition 4 relates the concavity of the distribution to its variance. For the Pareto distribution $\sigma(\theta) = -(\lambda+1)\frac{\theta}{\theta}$, where λ is the shape parameter, and for the exponential distribution $\sigma(\theta) = -\lambda$, where λ is the rate. Hence, credible pooling equilibria exist for the Pareto and exponential distributions when the shape (or rate) is greater than a specific level that depends on the model parameters. The variance for both of these distributions is a function of this λ parameter and in both cases it is decreasing in λ. The variance of the Pareto distribution is $[\lambda(\underline{\theta})^2]/[(\lambda-2)(\lambda-1)^2]$ which is infinite for $\lambda = 1$ or 2 and is otherwise decreasing in λ. The variance of the exponential distribution is $1/\lambda^2$ which is decreasing in λ. Combining these facts with propositions 3 and 4 and corollary 1 we can now state the following.

Corollary 2: *If assumptions 1−3 hold and the distribution is Pareto or exponential, then credible stereotype threat occurs when the variance of the distribution of types is below a critical level. The measure of credible pooling equilibrium expenditures for any set of fully, or partially, pooling productivities is larger if the variance of the distribution is smaller.*

Corollary 2 indicates that stereotype threat is more likely if there is little difference between the types. In this case the label is a useful heuristic device. On the other hand, if there is more dispersion in the types, then the label is less meaningful and self-fulfilling reputations are less likely to occur.

Although not covered by corollary 2 a similar result occurs for the uniform distribution. It is log concave with a weakly log-convex density, like the exponential and Pareto distributions, however, the uniform density is constant so that $\sigma(\theta) = 0$. Still, we can analyze the effect of the variance in the two productivity case. Let $v(h) = 2$, $v(l) = 1$, $c(q, s) = s/v(q)$, $I(\theta, q) = \theta[v(q) - 1]$, and $\theta \sim U[0, a]$. In this case $\theta^h = s^p/2$ and $w^p = 2\theta^h/a + (a - \theta^h)/a = 1 + s^p/2a$. A credible pooling equilibrium requires that $1 + s^p/2a - s^p/2 \geq 3/2$. Rearranging this expression and noting that $s^p < 1 = s^s$ yields $1 > s^p \geq a/(1 - a)$. Hence, credible pooling equilibria exist for $a < ½$. Furthermore, $\theta^h = s^p/2$ is bounded above by a, therefore, the interval of credible pooling signals is $2a > s^p \geq a/(1 - a)$. The size of this interval as a percentage all pooling signals is $[2a - a/(1 - a)]/a = 2 - 1/(1 - a)$, which is decreasing in a. If a is near ½, then the percent of credible s^p is small. If a is reduced to zero, then the percent of credible s^p is maximized.

Finally, note that the variance of this uniform distribution is $a^2/12$, so that the measure of credible pooling equilibria is decreasing in the variance of the distribution of types.

C. Counter-Stereotypical Behaviour.

We now consider some interesting properties of the set of credible pooling equilibria. Referring to figure 5 we see that for any group of pooling productivities there is a compact continuous set of education levels and resulting pooling wages such that the pooling equilibrium indifference curves (for all of the pooling productivities) through this wage lie above the separating equilibrium indifference curves for these same productivities. Note, as well, that for $s^p < \overline{s}^c(l,m,h)$ the separating indifference curve for a high productivity worker lies above that for a lower productivity worker. Hence, when starting at a credible (fully or partial) pooling equilibrium, if s^p drops below the lower bound of this credible set, then it is always the highest pooling productivity worker who would first choose to break the lower non-credible pooling equilibrium. This is the idea of counter-stereotypical behaviour.

Figure 5 goes about here.

The intuition behind counter-stereotypical behaviour can be seen as follows. Consider three different pooling equilibria indicated in figure 5. In the best equilibrium, marked with an "A" (and belonging to the "A" label), all three productivities pool together with a high level of tertiary education. The fully pooling equilibrium for the "B" label has less education, a lower wage, and lower expected productivity. The equilibrium for the "C" label has an even lower level of education. At this level of education the offered wage is low enough that the highest productivity worker would separate and a fully pooling equilibrium at credential level "C" is not credible. In the credible equilibrium the lower two productivities pool at credential level "C", with an even lower wage and expected average productivity and the highest productivity separates with a credential that is even higher than that for the "A" label.

Although label "B" is unequivocally in a worse equilibrium than label "A", the welfare comparison for label "C" is more complicated. The lower pooling productivities with label "C" are clearly worse off in the lower equilibrium, but the higher separating productivity does have a higher education and wage (but lower expected utility). The idea here is that moderately discriminatory beliefs may hurt a group, however, very discriminatory beliefs can serve as an impetus for the more capable in that group to distinguish themselves and separate. Hence, a higher percentage of this discriminated against group attains a high level of university education and this high level is larger than that obtained by a worker with a label that does not suffer any initial discrimination. In this way non-credible pooling beliefs generate counter-stereotypical behaviour.

In figure 5, in addition to the three indifference curves we show three pooling wage functions. The bottom wage function is when low and medium pool. The top two are when all three productivities pool. Consider first the topmost wage function. This wage function crosses the separating indifference curve for the high productivity at the point labeled on the s axis as $\underline{s}^c(l, m, h)$. All pooling education levels that are between this intersection and where this pooling wage function crosses the low productivity indifference curve are credible, therefore, stereotype threat with three productivities can occur. For all pooling education levels that are below $\underline{s}^c(l, m, h)$ the high productivity worker would separate. If the pooling education level was still greater than $\underline{s}^c(l, m)$, then pooling for the low and middle productivities would still be part of a credible equilibrium. In this way the stereotype threat remains while counter-stereotypical behaviour occurs simultaneously.

From proposition 4 we know that the middle pooling wage function shown (with a thinner line) in figure 5 could occur, if the investment cost function was relatively steeper in θ or if the distribution function was relatively flatter. A smaller measure of types, therefore, become high productivity for any university education level and the pooling wage function reflects their reduced number. What is interesting about this middle pooling wage function is that the minimum university education that is part of a credible three-way pooling equilibrium is greater than that for the higher pooling wage function. In this case, the high productivity is more likely to break a fully pooling equilibrium and there are a greater measure of credible equilibria whereby stereotype threat and counter-stereotypical behaviour occur simultaneously. In developing these ideas formally we define the maximal set of credible pooling productivities for any pooling expenditure s^p as the largest set of productivities, starting at q_0, that can be part of a credible pooling equilibrium for s^p. Formally, we write

$$\{ \overline{q}_J \left(s^p\right) \mid \text{the largest } q_J \text{ such that } Eu[\underline{\theta}, q_J, s^p, \boldsymbol{\rho}] \geq Eu[\underline{\theta}, q_J, s^{sJ}, \boldsymbol{\rho}, b_J = 1]\}. \tag{16}$$

With the definition given in equation 15 we can now state proposition 5 (the proof is in the appendix.)

Proposition 5: *If assumptions 1 – 3 are satisfied, then in any credible equilibrium the following hold.* (*i*) *For any maximal set of credible pooling productivities, where* $\overline{q}_J\left(s^p\right) < q_N$*, there is a continuum of the lowest types starting at* $\underline{\theta}$ *that choose productivities higher than* $\overline{q}_J$ *and separating credentials larger than* s^p. (*ii*) *This maximal* $\overline{q}_J$ *is lower if* $\sigma(\theta)$ *is larger and/or* $I_\theta(\cdot, q)$ *is larger.*

Proposition 5 complements proposition 4 in the following way. Proposition 4 shows that if the distribution function is steeper and/or the investment cost function is flatter, then the measure of credible pooling expenditures for any group of pooling productivities, and also the maximal set of credible pooling productivities number, are both larger. Proposition 5 gives a partial converse. If the distribution function is a little flatter and/or the investment cost function is a little steeper, then credible stereotype threat can exist along with credible counter-stereotypical behavior. We can put together the results of proposition 5 and corollary 2 as well as our uniform distribution example to note the following corollary 3.

Corollary 3. *If assumptions 1-3 are satisfied, and if the variance in investment costs are of an intermediate level, and/or if the distribution of types is exponential, Pareto, or uniform and the variance of types is of an intermediate level, then credible stereotype threat can occur along with stereotype threat. As either of these variances increase the maximal* $\overline{q}_J\left(s^p\right)$ *decreases and the extent of counter-stereotypical behavior increases. As either of these variances decrease the maximal* $\overline{q}_J\left(s^p\right)$ *increases.*

The two main results of the paper are the existence of stereotype threat as a credible equilibrium in a signalling model and how it may generate, and co-exist with, counter-stereotypical behavior. The variance of the distribution of types or of the investment costs (which are determined by the types) can determine the resulting equilibrium. If the variance is extremely small, then given our cost assumptions all of the types prefer the fully pooling equilibrium. As the variance is reduced the highest productivity workers prefer to separate from a lower pooling equilibrium, so that they engage in counter-stereotypical behavior while the lower productivities suffer from stereotype threat. Of course, if the variance is very large, then the only credible equilibrium is the lowest cost fully separating one.

D. Counter-Stereotypical behaviour and the Intuitive Criterion.

We now show that none of our results depends on our use of the less restrictive dominance refinement. In our model, as in a standard signalling model with more than two productivities, the intuitive criterion (and equilibrium dominance) does not eradicate all pooling equilibria. Hence, stereotype threat along with counter-stereotypical behaviour can exist in a signalling model when beliefs satisfy the intuitive criterion. In this section we demonstrate this claim as well as illustrate some potential differences that may arise when equilibrium beliefs are subject to this more restrictive refinement.

The intuitive criterion proceeds by first looking for signals that are equilibrium dominated for a particular productivity. That is, an alternative signal, s', is equilibrium dominated for productivity q' if the payoff at the proposed equilibrium is larger than the maximum payoff that could be obtained by s'. Note that the equilibrium payoff is generally higher than the minimum payoff considered in E2bd, so that equilibrium dominance will prune more signals for productivity q'. If s' is equilibrium dominated for q' we then eradicate q' as a possible sender of s' and consider the minimum payoff to s' if the receiver believes the sender is a remaining type (for whom s' is not equilibrium dominated). With only two productivities, if a signal, s', is equilibrium dominated for the low productivity worker, then beliefs must ascribe probability one to the event that it was sent by a high productivity worker and the minimum payoff to the sender of s' must reflect these beliefs. With two productivities this test amounts to looking for a signal that, even when believed to come from the high-productivity worker, is equilibrium dominated for the low-productivity worker, but not for the high-productivity worker. The single-crossing-property ensures that starting from any pooling equilibrium such a signal can always be found.

With three or more productivities, however, equilibrium dominance does not have as much bite. For example, suppose that the low and medium productivity workers pool at s^{plm}, but the high productivity worker separates at s^{sh}. The maximum response to a signal s' is that it came from a high and not just a medium productivity worker. In figure 6, consider the partial pooling equilibrium where the two dashed indifference curves cross a pooling wage function at s^{pg}. The intersection of the low productivity indifference curve (through the conjectured pooling equilibrium) with the full information wage for the high productivity worker yields the minimum signal that is equilibrium dominated for the low productivity: s^{0}. If the low productivity worker is pruned from the set of possible senders, then the minimum response to s^0 is that it came from a medium, but not a high, productivity worker. Drawing the medium productivity indifference curve though this minimum response to s^0 we see that s^0 is then dominated by the equilibrium payoff for the medium productivity as well. Hence, in this case, an equilibrium where the low and medium productivities pool (as in the previous section) and the high productivity separates satisfies the intuitive criterion.

Figure 6 goes about here.

We consider an additional lower partial pooling equilibrium in figure 6 (shown by the non-dashed indifference curves that cross at s^{pb}. It corresponds to the minimum partial pooling equilibrium that satisfies the dominance refinement. In particular, the utility level (and indifference curve) for the medium productivity in this pooling equilibrium is the same as that in the separating equilibrium (and its corresponding indifference curve). As seen in figure 6 the utility of the low productivity is higher in this lower pooling equilibrium and the minimum signal that is equilibrium dominated for low is slightly less than s^0, we will call it $s^{0'}$. Although not explicitly drawn in figure 6, it is easy to see that the indifference curve for the medium productivity that goes through the minimum response to this $s^{0'}$ (even after low is pruned from the possible senders) is less than medium would receive in the pooling equilibrium. The medium productivity could receive this same response by choosing the less costly separating credential, s^{sm}. Hence, in this example (which corresponds to the parameterization given in footnote 14) the intuitive criterion yields exactly the same set of partial pooling equilibria as the less restrictive dominance refinement.[25]

[25] A similar result occurs with the uniform distribution example first given in section IV.B. Adding a third productivity we have: $v(h) = 3$, $v(m) = 2$, $v(l) = 1$, $c(q, s) = s/v(q)$, $I(\theta, q) = \theta(v(q) - 1)^3$ and $\theta \sim U[0, a]$, where $a = ¼$. θ^h is defined by $3 - s^{sh}/3 - 8\theta^h = w^p - s^p/2 - \theta^h$ and $\theta^m = s^p/2$, where $s^{sh} = 3$ and $w^p = (a + \theta^m - 2\theta^h)/(a - \theta^h)$. Substitution yields $\theta^h = (7 + 2s^p)/56 \pm \sqrt{[(7 + 2s^p)^2/16 - 28(¼ - 3s^p/8)]}/14$. The smaller root of this quadratic substituted into w^p yields the pooling wage function, which crosses the medium separating indifference curve from below at .39. Naturally, θ^m is bounded above by $a = ¼$, therefore, s^p is bounded above by ½ in this example. Hence, all $s^p \in (.39, .5)$ can be part of a credible pooling equilibrium. At $s^p = 2/5$ we have $\theta^m = 1/5$ and $\theta^h = .0677679$. The pooling wage is $w^p = 1.72562$. The maximized utility for the low, medium and high workers are 1.32562, $1.52562 - \theta$, and $2 - 8\theta$, therefore, this equilibrium satisfies the dominance refinement. The minimum signal that is equilibrium dominated for the low worker is $s^0 = 1.67438$. If the low worker is removed as a possible sender of s^0, then the minimum wage to a sender of s^0 is $v^m = 2$ and the minimum payoff to the medium sender of s^0 is $1.16281 - \theta < 1.52562 - \theta$, therefore, s^0 is equilibrium dominated for the medium worker even after low has been omitted as a possible sender. Hence, this equilibrium satisfies the intuitive criterion and by continuity so do nearby pooling equilibria. In fact, for this simple example all pooling equilibria that satisfy our dominance refinement in E2bd also

The one main difference between our dominance refinement and the intuitive criterion is that with the latter the highest productivity, q_N, must separate. To speak of counter-stereotypical behavior we then need to see if, at least q_{N-1} would choose to separate, while q_{N-2} and q_{N-3} pool. Hence, under the intuitive criterion, stereotype threat occurring along with counter-stereotypical behavior requires at least four productivities whereas it requires only three under the dominance requirement. The dominance refinement allows for a less complex graphical presentation and for that reason we choose to develop the intuition with the less restrictive refinement.[26]

V. Conclusion

We introduce stereotype threat in a signalling model. The novel feature of our model is that we allow the worker to choose his productivity, but we tie this choice to a cost that is given by his unobservable type. It is this additional choice which generates a stereotype threat effect. The existence of multiple self-fulfilling stereotypes that satisfy a reasonable beliefs refinement is shown to be more likely if there is less variance in the distribution of types. We also show that a low endogenously correct stereotype forces a higher productivity worker with that stereotype to separate and, thereby, engage in counter-stereotypical behaviour. In this way a higher productivity worker from a discriminated against label overtakes a complacent pooling worker with a good stereotype label so that the good stereotype can generate a reputational Dutch disease.

The strategic response to discrimination that exists in our model is a novel explanation for why Canadian-born visible minorities receive more education than Canadian-born whites, as well as why intermediate and higher ability African-Americans receive more education than similar ability whites. The theory presented here also indicates that if the higher ability members of a label engage in counter-stereotypical behaviour, then the remaining partially-pooling group will have even lower average productivity. Our theory, therefore, predicts that a label with a lower stereotype would see greater income

satisfy equilibrium dominance.

[26] Although it does not affect the measure of credible partial pooling equilibrium credentials in our drawn example, the intuitive criterion may reduce them for other parameter values. Referring to figure 6 it is easy to see that s^0 is increasing in v^h. This observation suggests that if $v^h - v^m$ is small, then the measure of partial pooling equilibrium that satisfy the intuitive criterion may be smaller than those that satisfy the dominance refinement.

disparity. We leave the empirical verification of this newly discovered effect of statistical discrimination on income distribution for further research.

Whereas the reputational Dutch disease and the effect of discrimination on income distribution is unique to our framework, Lang and Manove (2011) also show that some members of a discriminated against label may choose extra education. An important difference between the models is that their results rely on a measurement error assumption. If they were to change their assumption on the location of the measurement error, then they would also change which part of the ability distribution obtains more education. The prediction in our model, on the other hand, is fixed because counter-stereotypical behaviour only occurs for the upper-intermediate and higher ability worker. Whereas our model is better able to fit the Canadian and African-American women data, it incorrectly predicts that lower intermediate ability African-American men would have less education than similar ability whites. We think it is possible to improve this aspect of our model by allowing the distribution of types to differ across labels.

An important extension to our model is, therefore, to allow the distribution of types to differ across labels. In particular, differences in school quality, neighborhood peer effects, or available parental resources may generate differences in the average costs of lifetime productivity investments (which are assumed to be related to choices made during the years of primary and secondary schooling) across labels. With this modification our model may more closely capture the empirical evidence on earnings, race, education and performance on the AFQT that is reported in Neil and Johnson (1996) and Lang and Manove (2011). An additional extension would allow for more than one partially pooling group so that there may be a low pooling group, a high pooling group, and a couple of very high separating productivities. This extension could prove significant if we also allow the composition (the number of pooling productivities) of these groups to differ across labels. In addition, combining differences in the type distributions with differences in the composition of the pooling groups may further improve the empirical relevance of our model. We leave these topics for further research.

Appendix

Proof of Proposition 1: By definition $Eu[\theta, q_k, s^{sk}, \boldsymbol{\rho}, b_k = 1] = Eu[\theta, q_{k-1}, s^{sk-1}, \boldsymbol{\rho}, b_{k-1} = 1]$ and, by assumptions 1 and 2, $s^{sk} > s^{sk-1}$, therefore, for all $s > s^{sk}$ it is true that $Eu[\theta, q_{k-1}, s, \boldsymbol{\rho}, b_k = 1] <$ $Eu[\theta, q_{k-1}, s^{sk-1}, \boldsymbol{\rho}, b_{k-1} = 1]$. Hence, all $s > s^{sk}$, are strictly dominated strategies for the q_{k-1} productivity worker for any posterior beliefs and a belief $\Pr(q_{k-1}|s > s^{sk}) > 0$ does not satisfy (E2bD). Furthermore, a posterior belief $\Pr(q_k|s) = 1$ for some $s < s^s$ causes the q_k productivity worker to deviate to that s. Hence, if there are posterior beliefs such that s^{sk} is not a dominated strategy for the q_k productivity worker, then there is an equilibrium where credible posterior beliefs are $\Pr(q_k|s < s^{sk}) = 0$ and $\Pr(q_k|s^{sk} \leq s < s^{sk+1}) = 1$.

We now show that if the posterior beliefs are $\Pr(q_k|s < s^{sk}) = 0$ and $\Pr(q_k|s^{sk} \leq s < s^{sk+1}) = 1$, then a q_k productivity worker's unique best response is s^{sk}.

We need to show that $Eu[\theta, q_k, s^{sk}, \boldsymbol{\rho}, b_k = 1] > Eu[\theta, q_k, s^{sk-1}, \boldsymbol{\rho}, b_{k-1} = 1]$, which is equivalent to $Eu[\theta, q_k, s^{sk}, \boldsymbol{\rho}, b_k{=}1] - Eu[\theta, q_k, s^{sk-1}, \boldsymbol{\rho}, b_{k-1}{=}1] > 0 = Eu[\theta, q_{k-1}, s^{sk}, \boldsymbol{\rho}, b_k{=}1] - Eu[\theta, q_{k-1}, s^{sk-1}, \boldsymbol{\rho}, b_{k-1}{=}1]$
$\Leftrightarrow v(q_k) - C(\theta, q_k, s^{sk}) - v(q_{k-1}) + C(\theta, q_k, s^{sk-1}) > v(q_k) - C(\theta, q_{k-1}, s^{sk}) - v(q_{k-1}) + C(\theta, q_{k-1}, s^{sk-1})$
$\Leftrightarrow C(\theta, q_{k-1}, s^{sk}) - C(\theta, q_k, s^{sk}) > C(\theta, q_{k-1}, s^{sk-1}) - C(\theta, q_k, s^{sk-1})$, which is true by assumption 1.

Finally note that that $Eu[\theta^{sk}, q_k, s^{sk}, \boldsymbol{\rho}, b_k = 1] = Eu[\theta^{sk}, q_{k-1}, s^{sk-1}, \boldsymbol{\rho}, b_{k-1} = 1]$ and for all $\theta < \theta^{sk}$, $Eu[\theta, q_k, s^{sk}, \boldsymbol{\rho}, b_k = 1] > Eu[\theta, q_{k-1}, s^{sk-1}, \boldsymbol{\rho}, b_{k-1} = 1]$; therefore, all $\theta \in [\underline{\theta}, \theta^{sN}]$ choose $\{q_N, s^{sN}\}$ all $\theta \in (\theta^{sk+1}, \theta^{sk}]$, choose $\{q_k, s^{sk}\}$, and all $\theta > \theta^{s1}$, choose $\{q_0, s^{0*}\}$.

Now rewriting the equation for the indifferent type and using the definition of a separating signal yields $Eu[\theta^{sk}, q_k, s^{sk}, \boldsymbol{\rho}, b_k = 1] = Eu[\theta^{sk}, q_{k-1}, s^{sk-1}, \boldsymbol{\rho}, b_{k-1} = 1] = Eu[\theta^{sk}, q_{k-1}, s^{sk}, \boldsymbol{\rho}, b_k = 1]$, which implies that $v(q_k) - C(\theta^{sk}, q_k, s^{sk}) = v(q_k) + C(\theta^{sk}, q_{k-1}, s^{sk})$ which, using assumption 1, can rewritten as:
$\mathrm{I}(\theta^{sk}, q_k) - \mathrm{I}(\theta^{sk}, q_{k-1}) = g(\theta^{sk})[d(q_k) - d(q_{k-1})] = c(q_{k-1}, s^{sk}) - c(q_k, s^{sk})$. Now, $g(\theta)$ is continuous and strictly increasing, therefore, it has a continuous and strictly increasing inverse and the critical type for each quality is given as: $\theta^{sk} = g^{-1}\left(\dfrac{c\left(q_{k-1}, s^{sk}\right) - c(q_k, s^{sk})}{d\left(q_k\right) - d\left(q_{k-1}\right)}\right)$. By assumption 1, $c(q_{k-1}, s^{sk}) - c(q_k, s^{sk})$ is positive and decreasing as productivity further increases and $d(q_k) - d(q_{k-1})$ is positive and non-decreasing in further productivity increases. Hence $\theta^{sk} > \theta^{sk+1}$ and because $F(\theta)$ is increasing and continuous we have that $\rho^k = F(\theta^{sk}(\boldsymbol{\rho})) - F(\theta^{sk+1}(\boldsymbol{\rho}))$ for all q_k. □

Proof of Proposition 3: Rewriting equation (4) for the lowest separating credential for productivity q_{J+1} (when productivities q_0 through q_J pool) yields:

$$Eu[\theta, q_J, s^{sJ+1}, \boldsymbol{\rho}, b_{J+1} = 1] - Eu[\theta, q_J, s^{pJ}, \boldsymbol{\rho}] = 0. \tag{A1}$$

Hence s^{sJ+1} is determined in relation to s^{pJ}. Proceeding as in proposition 1 yields s^{sJ+t} for $t \in [1,\ldots, N-J]$.

Now rewriting the equation for the indifferent type and using (A1) yields
$Eu[\theta^{J+1}, q_{J+1}, s^{sJ+1}, \boldsymbol{\rho}, b_{J+1} = 1] = Eu[\theta^{J+1}, q_J, s^{pJ}, \boldsymbol{\rho}] = Eu[\theta^{J+1}, q_J, s^{sJ+1}, \boldsymbol{\rho}, b_{J+1} = 1]$ which implies that $v(q_{J+1}) - C(\theta^{J+1}, q_{J+1}, s^{sJ+1}) = v(q_{J+1}) - C(\theta^{J+1}, q_J, s^{sJ+1})$ which can rewritten as:

$\theta^{J+1} = g^{-1}\left(\dfrac{c\left(q_J, s^{sJ+1}\right) - c(q_{J+1}, s^{sJ+1})}{d\left(q_{J+1}\right) - d\left(q_J\right)}\right)$. Hence, using the same argument as in the proof of proposition 2, θ^{J+1} is strictly increasing in s^{sJ+1} and using the same argument as in the proof of proposition 1, θ^{J+t} is decreasing in t.

We now show that the partial pooling wage is increasing in s^{pJ}. In a partial-pooling equilibrium the expected productivity and the pooling wage is determined as a conditional probability:

$$w^p(s^{pJ}, \boldsymbol{\rho}) = \frac{\sum_{k=0}^{J} \rho^k \cdot v(q_k)}{1-\sum_{k=J+1}^{N} \rho^k} = \frac{\sum_{k=0}^{J}[F(\theta^k) - F(\theta^{k+1})]v(q_k)}{1-\sum_{k=J+1}^{N}[F(\theta^k) - F(\theta^{k+1})]} = \frac{\sum_{k=0}^{J} v(q_0) + F(\theta^{k+1})[v(q_{k+1}) - v\left(q_k\right)]}{1-F(\theta^{J+1})}$$

Its derivative with respect to s^{pJ} is:

$$\frac{\partial w^p(s^{pJ}, \boldsymbol{\rho})}{\partial s^{pJ}} =$$

$$\frac{\sum_{k=0}^{J} F_\theta(\theta^{k+1})\delta_s\left(\cdot, q_{k+1}\right)[v(q_{k+1}) - v\left(q_k\right)]}{1-F(\theta^{J+1})} + \frac{\left(v(q_0) + \sum_{k=0}^{J} F(\theta^{k+1})[v(q_{k+1}) - v\left(q_k\right)]\right) F_\theta(\theta^{J+1}) \frac{\partial \theta^{J+1}}{\partial s^{ppJ}}}{[1-F(\theta^{J+1})]^2}. \quad \text{(A2)}$$

The proof that this derivative is positive is argued by contradiction. To this end, suppose that $\dfrac{\partial w^p}{\partial s^{pJ}} \leq 0$.

Because $c_s(q_k, \cdot) > 0$, this implies that $\dfrac{\partial w^p}{\partial s^{pJ}} < c_s(q_J, s^{ppJ})$. Taking the total differential of (A1) yields:

$$-\frac{\partial c(q_J, s^{sJ+1})}{\partial s^{sJ+1}} ds^{sJ+1} - \left[\frac{\partial w^p}{\partial s^{pJ}} - \frac{\partial c(q_J, s^{pJ})}{\partial s^{pJ}}\right] ds^{pJ} = 0, \text{ so that}$$

$$\frac{ds^{sJ+1}}{ds^{pJ}} = -\frac{\left[\dfrac{\partial w^p}{\partial s^{pJ}} - \dfrac{\partial c(q_J, s^{pJ})}{\partial s^{pJ}}\right]}{\dfrac{\partial c(q_J, s^{sJ+1})}{\partial s^{sJ+1}}}.$$

Hence, if $\dfrac{\partial w^p}{\partial s^{pJ}} \leq 0$, then $\dfrac{ds^{sJ+1}}{ds^{pJ}} > 0$ and using that $\dfrac{\partial \theta^{J+1}}{\partial s^{sJ+1}} >0$ implies that $\dfrac{\partial \theta^{J+1}}{\partial s^{pJ}} >0$. Now the first term on the right hand side of (A2) is positive (as given in the proof of proposition 2) and the sign of the second term depends on the sign of $\dfrac{\partial \theta^{J+1}}{\partial s^{pJ}}$. Therefore, if $\dfrac{\partial w^p}{\partial s^{pJ}} \leq 0$, then $\dfrac{\partial w^p}{\partial s^{pJ}} > 0$ which develops the necessary contradiction. Finally, note that the group average stereotype $\mathrm{E}q(s^{pJ}, \boldsymbol{\rho}) = \dfrac{\sum_{k=0}^{J} \rho^k \cdot q_k}{1-\sum_{k=J+1}^{N} \rho^k} = \dfrac{\sum_{k=0}^{J} q_0 + F(\theta^{k+1})[q_{k+1} - q_k]}{1-F(\theta^{J+1})}$ and the sign of $\dfrac{\partial Eq(s^{pJ}, \boldsymbol{\rho})}{\partial s^{pJ}}$ = the sign of $\dfrac{\partial w^p(s^{pJ}, \boldsymbol{\rho})}{\partial s^{pJ}}$. □

Proof of Proposition 4: (i) In a credible pooling equilibria no pooling productivity would deviate to the separating signal. Given that the productivity choices have been made and the investment cost have been sunk this implies that $w^p(s^p, \boldsymbol{\rho}) - c(q_k, s^p) \geq \text{Max } \{v(q_k) - c(q_k, s^{sk}), v(q_0) - c(q_0, s^{0*})\}$ or that:

$$w^p(s^p, \boldsymbol{\rho}) \geq c(q_k, s^p) + \text{Max } \{v(q_k) - c(q_k, s^{sk}), v(q_0) - c(q_0, s^{0*})\}. \tag{A2}$$

Hence, $w^p(s^p, \boldsymbol{\rho})$ must lie above the separating equilibrium indifference curves. From assumption 1 and proposition 1 we know that $v(q_k) - c(q_k, s^{sk}) > v(q_k) - c(q_{k-1}, s^{sk}) = v(q_{k-1}) - c(q_{k-1}, s^{sk-1}) > v(q_0) - c(q_0, s^{0*})$ for $k > 1$ so that, when $s^p = 0$, the separating equilibrium indifference curve for a higher productivity lies above that for a lower productivity. From assumption 1 and the fact that $s^{sk} > s^{sk-1} > s^{0*}$ this ordering is eventually reversed as s^p increases. Hence, for any partial or fully pooling equilibrium where productivities q_0 through q_k pool there exists a critical $\overline{s}_k$ such that for $s^p < \overline{s}_k$ the indifference curve for productivity q_k is above the others (in w^p, s^p space) and for $s^p > \overline{s}_k$ the indifference curve for productivity q_0 is topmost. For credibility of any pooling equilibrium we, therefore, only need to show that $w^p(s^p, \boldsymbol{\rho})$ lies above these two indifference curves.

From assumption 3, if, for $k \geq 1$, $s^p < s^{sk}$ and if, for $j < k$, ρ^j is close to zero, then at least $\underline{\theta}$ prefers the pooling equilibrium. If, in addition, $\sigma(\theta)$ is sufficiently negative, then the entire distribution converges to a mass point at $\underline{\theta}$, the correct stereotype for ρ^j approaches zero and all types prefer the pooling equilibrium. For any two distributions, if $\sigma_1(\theta) < \sigma_2(\theta)$ for all θ, then $F_1(\cdot)$ first-order stochastically dominates $F_2(\cdot)$, therefore, for any s^p we have that $F(\delta(s^p, q_k))$ is increasing in the degree of concavity of the distribution. Finally, note from equation (14) that $w^p(s^p, \boldsymbol{\rho})$ is increasing faster in s^p when $\sigma(\theta)$ is lower. Hence, there exists a $\lambda^* \geq 0$ such that if $\sigma(\underline{\theta}) \leq -\lambda^*$, then $F(\delta(s^p, q_k))$ is sufficiently large and equation (A2) holds as a strict inequality for some $q_k > q_0$ and some $\theta^k > \underline{\theta}$.

(*ii*) From assumption 3 note that $Eu[\underline{\theta}, q_1, s^{s1}, \boldsymbol{\rho}, b_1 = 1] > Eu[\underline{\theta}, q_0, s^{0*}, \boldsymbol{\rho}, b_0 = 1]$ so that at least $\underline{\theta}$ would break the most pessimistic pooling equilibrium.

(*iii*) If $w^p(s^p, \boldsymbol{\rho})$ is steeper in s^p, then, it first crosses each separating equilibrium indifference curve at a lower s^p and its second cross is at a higher s^p. Hence, if $w^p(s^p, \boldsymbol{\rho})$ is steeper in s^p it crosses the separating indifference curve for the highest pooling productivity at a lower s^p, the lowest pooling productivity at a higher s^p, and it crosses more pooling productivities. From equation (14) the derivative of the wage function with respect to s^p is $\dfrac{\partial w^p(s^p, \boldsymbol{\rho})}{\partial s^p} = \sum_{k=0}^{N-1} F_\theta(\theta^{k+1})\delta_s(\cdot, q_{k+1})[v(q_{k+1}) - v(q_k)]$, therefore, the pooling wage function is increasing faster in s^p when $F_\theta(\theta^k)$ is larger and from part (*i*) this term is larger when $\sigma(\theta)$ is smaller.

(*iv*) From equation (14), the pooling wage function is increasing faster in s^p when $\delta_s(\cdot, q_{k+1})$ is larger, and from equation (8), $\delta(s^p, q_k) = g^{-1}\left(\dfrac{c(q_{k-1}, s^p) - c(q_k, s^p)}{d(q_k) - d(q_{k-1})}\right)$ is larger when $g^{-1}(\cdot)$ is steeper, and $g^{-1}(\cdot)$ is steeper when $g(\theta)$ is flatter or when $I_\theta(\cdot, q)$ is smaller. □

Proof of Proposition 5: (*i*) If $\overline{q}_J\left(s^p\right) < q_N$, then $Eu[\underline{\theta}, q_{J+1}, s^{J+1}, \boldsymbol{\rho}, b_{J+1} = 1]\} > Eu[\underline{\theta}, q_{J+1}, s^p, \boldsymbol{\rho}]$. Hence, $s^{sJ+1} > s^p$. Now if $\underline{\theta}$ prefers choosing a productivity greater than $\overline{q}_J\left(s^p\right)$, then so do nearby θ. From the construction of the critical types given in equation (3), there is a θ^{J+1} who is just indifferent between $\{q_{J+1}, s^{sJ+1}\}$ and $\{\overline{q}_J\left(s^p\right), s^p\}$ and all $\theta < \theta^{J+1}$ strictly prefer $\{q_{J+1}, s^{J+1}\}$.

(*ii*) This part is the converse of part (*iv*) of proposition 4 which showed that the pooling wage function is steeper and, therefore, the maximal $\overline{q}_J\left(s^p\right)$ is higher if $\sigma(\theta)$ or $I_\theta(\cdot, q)$ are smaller. □

REFERENCES

Aigner, D. and G. Cain. 1977. "Statistical Theories of Discrimination in the Labor Market," *Industrial and labor Relations Review,* 30: 175-187.

Arcidiacono, P., P. Bayer and A. Hizmo. 2010. "Beyond Signaling and Human Capital: Education and the Revelation of Ability," *American Economic Journal: Applied Economics*, 2(4): 76-104.

Arrow, K. J. 1973. "The Theory of Discrimination," in O. Ashenfelter and A. Rees, eds., *Discrimination in Labor Markets*. Princeton, NJ: Princeton University Press.

Axtell, R. 2001. "Zipf Distribution of US Firm Sizes," *Science*, 293(7): 1818-1820.

Bagnoli, M. and T. Bergstrom. 2005. "Log-concave Probability and its Applications," *Economic Theory,* 26, 445-469.

Banks, J. and J. Sobel. 1987. "Equilibrium Selection in Signaling Games," *Econometrica,* 55: 647-662.

Bertrand, M. and S. Mullainathan. 2004. "Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination," *American Economic Review,* 94(4): 991-1013.

Bidner, C. 2012. "A Spillover-Based Theory of Credentialism," Working paper, University of New South Wales.

Cho, I. and D. Kreps. 1987. "Signaling Games and Stable Equilibria," *Quarterly Journal of Economics,* 102: 179-221.

Coate, S. and G. Loury. 1993. "Will Affirmative Action Policies Eliminate Negative Stereotypes," *American Economic Review,* 83: 1220-1240.

Coate, S. and S. Tennyson. 1992. "Labor Market Discrimination, Imperfect Information and Self Employment," *Oxford Economic Papers,* 44: 272-288.

Dvir, T., D. Eden and M. Banjo. 1995. "Self-fulfilling Prophecy and Gender: Can Women be Pygmalion and Galatea?" *Journal of Applied Psychology*, 80: 253-270.

Eden, D. 1992. "Leadership and Expectations: Pygmalion Effects and Other Self-fulfilling Prophecies in Organizations," *Leadership Quarterly*, 3: 271-305.

Feltovich, N., R. Harbaugh and T. To. 2002. "Too Cool for School? Signaling and CounterSignaling," *The Rand Journal of Economics*, 33(4): 630-649.

Ferrer, A. and W.C. Riddell. 2002. "The Role of Credentials in the Canadian Labour Market," *Canadian Journal of Economics*, 35(4): 879-905.

Ferrer, A. and W.C. Riddell. 2008. "Education, Credentials, and Immigrant Earnings," *Canadian Journal of Economics*, 41(1): 186-216.

Fryer, R. 2007. "Belief Flipping in a Dynamic Model of Statistical Discrimination," *Journal of Public Economics,* 91(5-6): 823-1230

Fudenberg, D. and J. Tirole. 1991. *Game Theory.* MIT Press, Cambridge, MA.

Galarza, F. and G. Yamada. 2012. "Labor Market Discrimination in Lima, Peru: Evidence from a Field Experiment," Working Paper, Universidad Del Pacífico.

Good, C. 2008. "Problems in the Pipeline: Stereotype Threat and Women's Achievement in High-level Math Courses," *Journal of Applied Developmental Psychology*, 29(1): 17-28.

Gupta, V., D. Turban and N. Bhawe. 2008. "The Effect of Gender Stereotype Activation on Entrepreneurial Intentions," *Journal of Applied Psychology*, 93:5.

Grossman, S. and M. Perry. 1986. "Perfect Sequential Equilibrium," *Journal of Economic Theory,* 39: 97-119.

Hanna, R. and L. Linden. 2012. "Discrimination in Grading," *American Economic Journal: Economic Policy,* 4(4): 146-168.

Kohlberg E. and J-F Mertens. 1986. "On the Strategic Stability of Equilibria," *Econometrica,* 54: 1003 -1038.

Kreps, D. and R. Wilson. 1982. "Sequential Equilibria," *Econometrica,* 50: 863-894.

Kiefer, A. 2007. "Implicit Stereotypes and Women's Math Performance: How Implicit Gender-math Stereotypes Influence Women's Susceptibility to Stereotype Threat," *Journal of Experimental Social Psychology,* 43(5): 825-832.

Lang, K. and J. Lehman. 2012. "Racial Discrimination in the Labor Market," *Journal of Economic Literature*, 50(4): 959-1006.

Lang, K. and M. Manove. 2011. "Education and Labor-Market Discrimination," *American Economic Review,* 101: 1467-1496.

Lange, F. 2007. "The Speed of Employer Learning," *Journal of Labor Economics*, 25(1): 1-36.

Lundberg, S. and R. Startz. 1983. "Private Discrimination and Social Intervention in Competitive Labor Markets," *American Economic Review,* 73: 340-347.

Mailath, G., M. Okuno-Fujiwara and A. Postlewaite. 1993. "Belief Based Refinements in Signaling Games," *Journal of Economic Theory,* 60: 241-276.

Mas-Colell, A., Whinston, M. and J. Green. 1995. *Microeconomic Theory.* Oxford University Press, New York.

Milgrom, P. and S. Oster. 1987. "Job Discrimination, Market Forces and the Invisibility Hypothesis," *Quarterly Journal of Economics,* 102: 453-476.

Neal, D. and W. Johnson. 1996. "The Role of Premarket Factors in Black-White Wage Differences," *Journal of Political Economy,* 104: 869–95.

Phelps, E. 1972. "The Statistical Theory of Racism and Sexism," *American Economic Review,* 62: 659-661.

Riley, J. 1979. "Informational Equilibria," *Econometrica,* 47: 331-360.

Rosenthal, R. and L. Jacobson. 1968. *Pygmalion in the Classroom: Teacher Expectation and Pupil's Intellectual Development.* Holt, Rinehart and Winston.

Salisbury, L. 2012. "Women's Income and Marriage Markets in the United States: Evidence from the Civil War Pension," Working Paper, Boston University, 2012.

Skuterud, M. 2010. "The Visible Minority Earnings Gap Across Generations of Canadians," *Canadian Journal of Economics*, 43(3): 860-881.

Spence, M. 1973. "Job Market Signaling," *Quarterly Journal of Economics*, 87: 355-374.

Steele C. and J. Aronson. 1995. "Stereotype Threat and the Intellectual Test Performance of African Americans," *Journal of Personal and Social Psychology,* 69(5): 797–81

Wilson, W. 1996. *When Work Disappears: The World of the New Urban Poor.* New York: Alfred Knopf.

Figure 1: Separating Equilibrium and the Critical Types

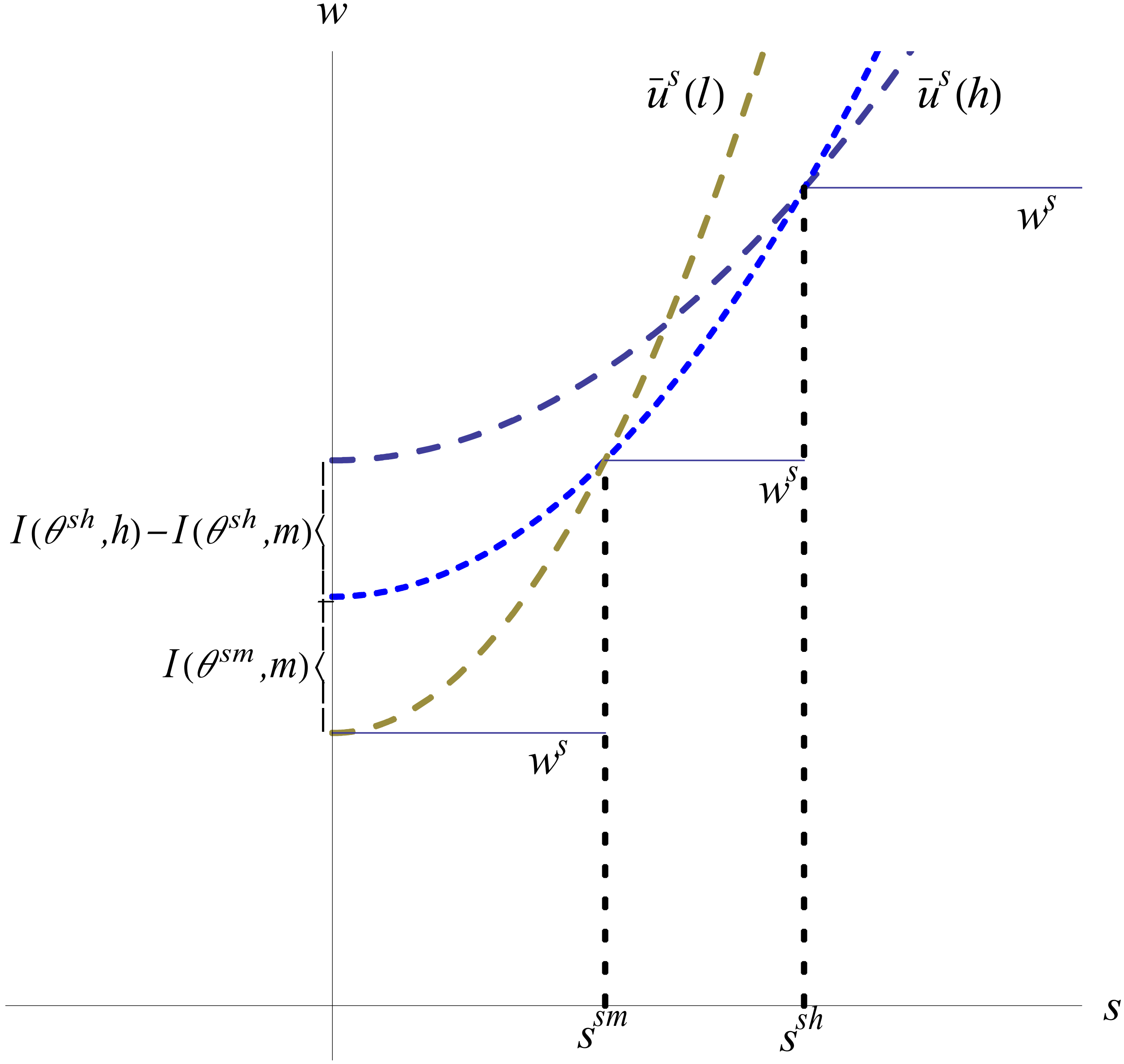

Figure 2: The Critical Type for Two Pooling Equilibria

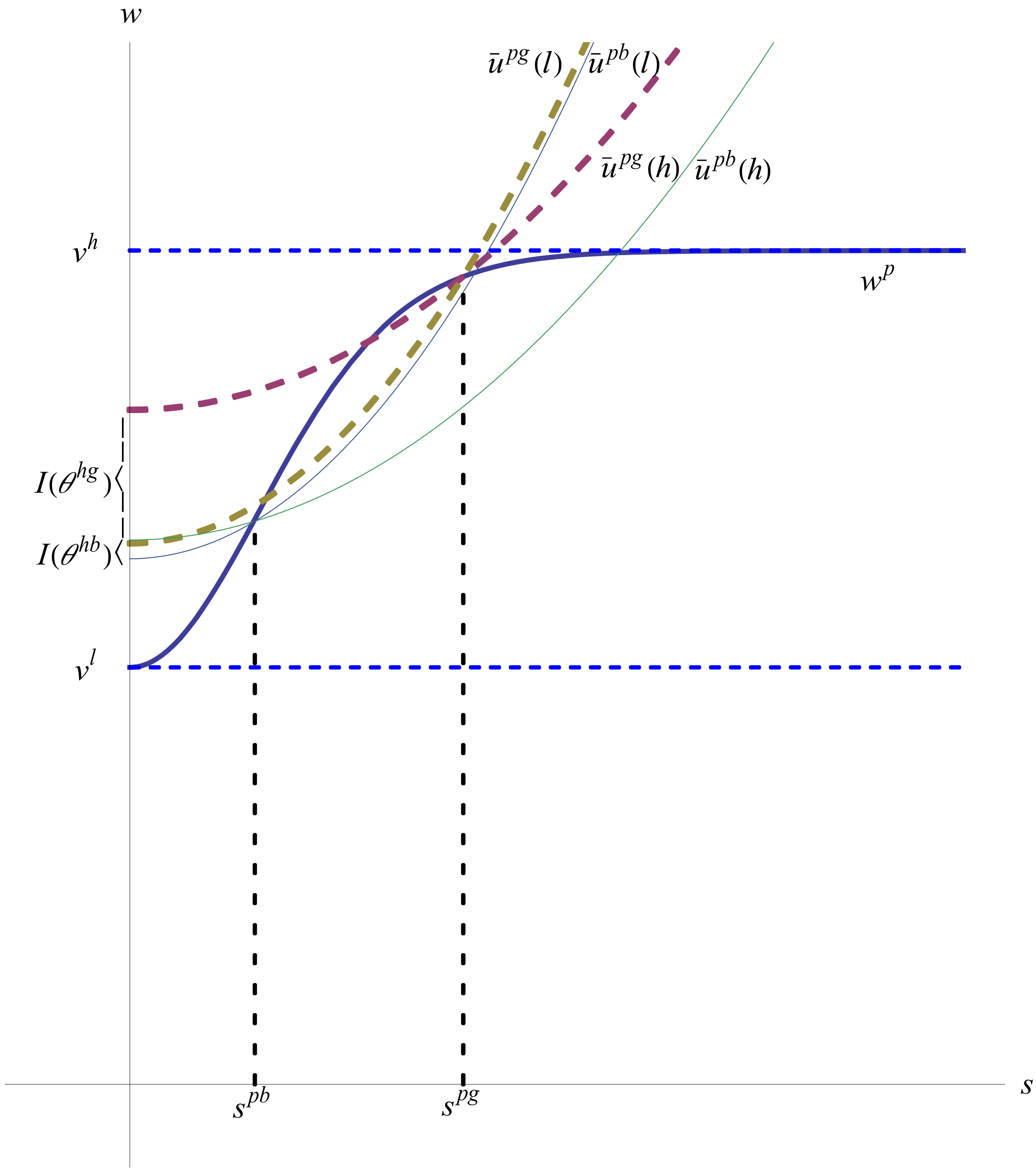

Figure 3: Credible Stereotype Threat

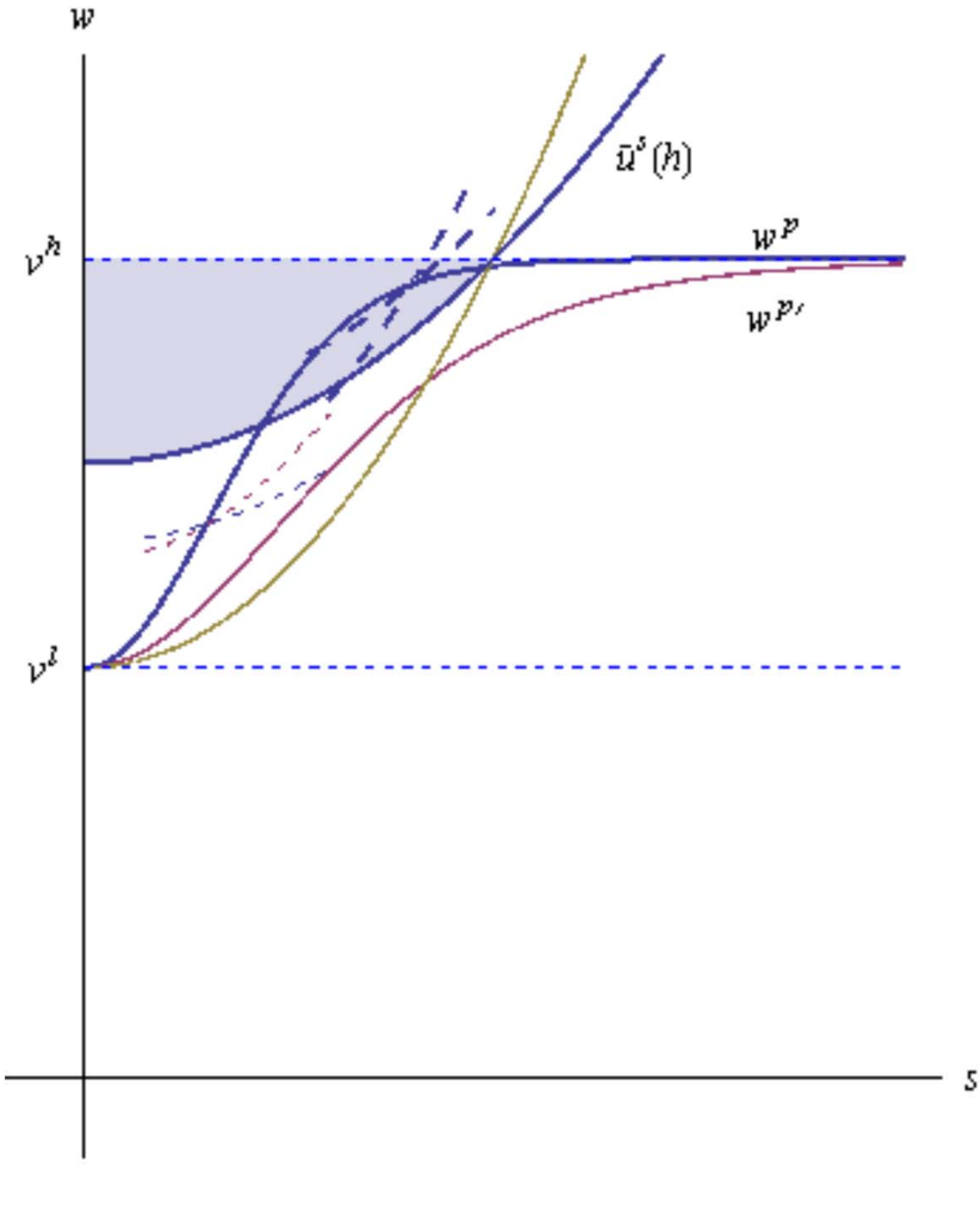

Figure 4: Credible Stereotype Threat with 3 Productivities

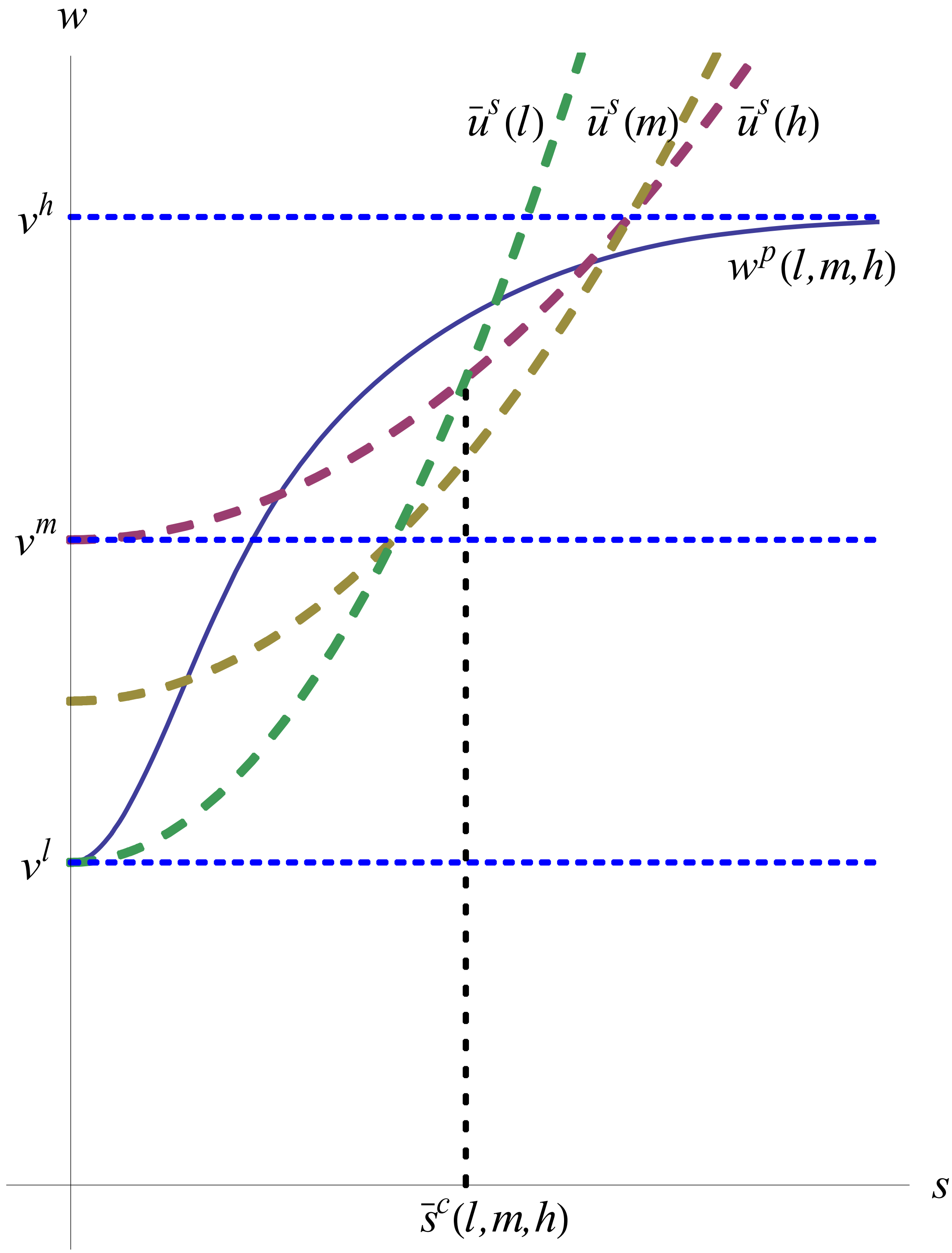

Figure 5: Stereotype Threat and Counter–Stereotypical Behaviour

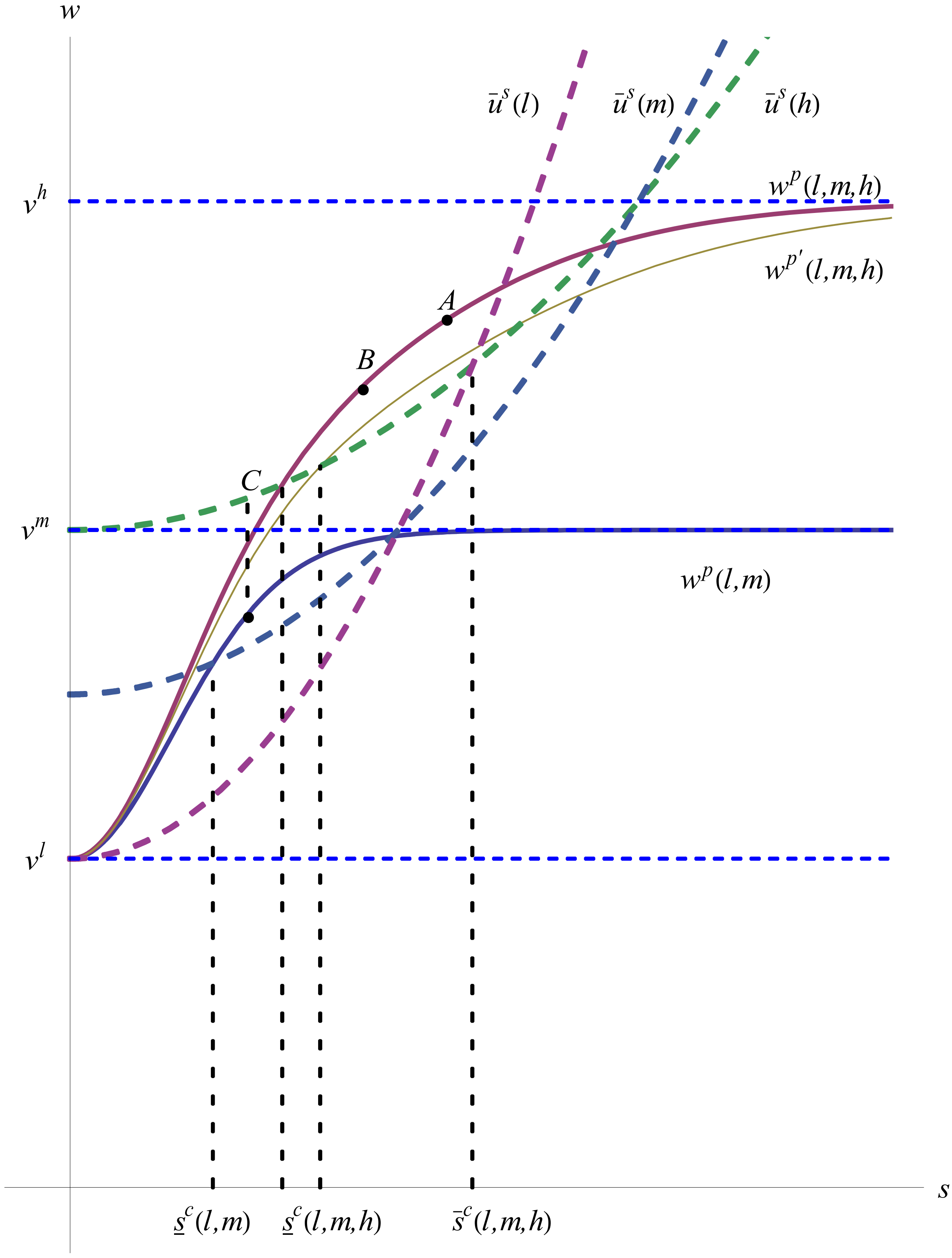

Figure 6: Stereotype Threat and the Intuitive Criterion